

OM of: US-09-303-518d-463 Lo: SwissProt\_40:\* out\_format : pfs

Date: Jun 30, 2002 8:36 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame\_n2p.model -DEV=xlh  
-Q/cgn2.1/USPTO\_spool/US09303518/runat\_28062002\_142714\_4362/app\_query.fasta\_1.23501  
-DB=SwissProt\_40 -OPMT=fastan -SUFFIX=esp -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-GAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -DEPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=100 -DOALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFWT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09303518 @CIGN1\_1440 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-463

Query length: 1887

Database: SwissProt\_40.\*

Database sequences: 105224

Database length: 38719550

Search time (sec): 217.960000

score.list:

Sequence	Strd Orig	zScore	Escore	Len
SwissProt_40:MUC2_HUMAN	165.50	160.64	0.0058	5179
SwissProt_40:FYB_MOUSE	160.50	171.44	0.0091	819
SwissProt_40:DANA_YEAST	153.50	160.49	0.0262	1161
SwissProt_40:AMVH_YEAST	146.00	150.64	0.0788	1367
SwissProt_40:NKCR_MOUSE	143.50	147.29	0.1139	1453
SwissProt_40:YAVI_SCHPO	143.50	145.42	0.1173	1794
SwissProt_40:KP38_DROME	141.00	148.25	0.1537	952
SwissProt_40:IGY3_HALSQ	139.00	152.94	0.1835	437
SwissProt_40:DAB_DROME	138.50	137.18	0.2510	2411
SwissProt_40:PODX_RABIT	138.00	149.76	0.2190	551
SwissProt_40:M4P4_MOUSE	138.00	143.40	0.2422	1125
SwissProt_40:YWR1_CAEEL	137.50	142.10	0.2633	1223
SwissProt_40:VRP1_YEAST	137.00	145.13	0.2673	817
SwissProt_40:VGLX_HSVB	136.50	144.79	0.2862	797
SwissProt_40:AMVB_HAEVE	133.00	137.26	0.5012	1196
SwissProt_40:DRPL_HUMAN	132.00	136.22	0.5779	1185
SwissProt_40:GTFB_STRMU	129.50	131.46	0.8538	1476
SwissProt_40:VGP3_EBVA8	129.00	135.44	0.8537	886
SwissProt_40:VE2_HPV08	128.50	140.01	0.8457	498
SwissProt_40:DRPL_RAT	126.50	130.07	1.27	1183
SwissProt_40:AGAL_YEAST	125.00	132.74	1.47	725
SwissProt_40:HIG_DROME	124.50	129.71	1.65	958
SwissProt_40:YUQ3_CAEEL	124.50	127.33	1.71	1251
SwissProt_40:V70K_TYMCV	124.00	132.90	1.67	628
SwissProt_40:Y58A_CAEEL	124.00	130.79	1.73	796
SwissProt_40:VGP3_EBV	123.50	129.07	1.89	907
SwissProt_40:EGRI_BRARE	123.00	133.62	1.87	511
SwissProt_40:A180_MOUSE	123.00	128.57	2.03	901
SwissProt_40:MGPC_MYCGE	123.00	127.19	2.07	1052
SwissProt_40:GTFC_STRMU	123.00	124.81	2.15	1375
SwissProt_40:SMFI_HUMAN	123.00	121.92	2.25	1902
SwissProt_40:V70K_TYMCV	122.50	131.22	2.07	628
SwissProt_40:NCR1_HUMAN	121.50	118.02	2.89	2440
SwissProt_40:PSC_DROME	121.00	121.20	2.93	1603
SwissProt_40:N214_HUMAN	120.50	118.28	3.27	2090
SwissProt_40:ALP_ECAPC	120.00	132.39	2.78	402
SwissProt_40:MUC4_HUMAN	120.00	128.68	2.95	610
SwissProt_40:DAGI_MOUSE	119.50	124.73	3.35	893
SwissProt_40:CAIE_CHICK	119.50	118.07	3.72	1888
SwissProt_40:PIP_LACLC	119.50	118.00	3.73	1902

SwissProt_40:CCAA_MOUSE	119.50	116.85	3.79
SwissProt_40:ITAA_HUMAN	119.00	122.83	3.67
SwissProt_40:MAP4_BOVIN	118.50	121.98	3.97
SwissProt_40:P2P_LACLC	118.50	116.88	4.30
SwissProt_40:CCAA_RAT	118.50	115.54	4.39
SwissProt_40:CGGT_BACCS	118.00	124.99	4.03
SwissProt_40:MAP4_HUMAN	118.00	120.78	4.31
SwissProt_40:A180_RAT	117.50	122.27	4.48
SwissProt_40:LGI_WAIZE	117.00	129.09	4.28
SwissProt_40:STUA_EMENI	117.00	125.61	4.52
SwissProt_40:V70K_TYMCV	117.00	125.06	4.56
SwissProt_40:IGA0_HAEIN	117.00	116.23	5.25
SwissProt_40:IGA2_HAEIN	117.00	116.19	5.25
SwissProt_40:M115_CAEEL	116.50	119.67	5.29
SwissProt_40:YM96_YEAST	116.00	118.63	5.73
SwissProt_40:IGA4_HAEIN	116.00	114.33	6.14
SwissProt_40:GTF2_STRDO	115.50	115.10	6.46
SwissProt_40:P3P_LACLC	114.50	112.40	7.64
SwissProt_40:YHC8_YEAST	114.00	122.03	6.99
SwissProt_40:EGRI_HUMAN	113.50	122.43	7.39
SwissProt_40:MUC1_XENLA	113.50	120.67	7.60
SwissProt_40:GBF_DICDI	113.50	120.07	7.68
SwissProt_40:YQ36_CAEEL	113.50	117.33	8.02
SwissProt_40:YQ3F_YEAST	113.50	113.74	8.49
SwissProt_40:ZEP1_MOUSE	113.50	108.20	9.27
SwissProt_40:ICP0_HSVBK	113.00	119.92	8.19
SwissProt_40:SON_MOUSE	113.00	108.63	9.80
SwissProt_40:MUC1_MESAU	112.50	119.36	8.80
SwissProt_40:SNP_DROME	112.50	118.10	8.98
SwissProt_40:FLO5_YEAST	112.50	115.23	9.40
SwissProt_40:TCOF_HUMAN	112.50	112.81	9.77
SwissProt_40:FLO1_YEAST	112.50	112.05	9.89
SwissProt_40:NGR2_HUMAN	112.50	107.66	10.60
SwissProt_40:ICP0_HSVBK	112.00	118.80	9.46
SwissProt_40:YAMB_CAEEL	112.00	116.08	9.88
SwissProt_40:NIA_SPIOL	112.00	116.00	9.89
SwissProt_40:PGCA_RAT	112.00	108.61	11.12
SwissProt_40:SON_HUMAN	112.00	107.43	11.33
SwissProt_40:WETA_PENCH	111.50	120.94	9.74
SwissProt_40:VG37_BPARI	111.50	113.88	10.89
SwissProt_40:IGA_NEIGO	111.50	110.96	11.41
SwissProt_40:YEW2_YEAST	111.50	109.76	11.63
SwissProt_40:NCAL_XENLA	111.00	113.44	11.68
SwissProt_40:FURC_DROME	111.00	113.34	11.70
SwissProt_40:YAG3_YEAST	111.00	111.71	12.01
SwissProt_40:TUSP_MOUSE	111.00	110.31	12.28
SwissProt_40:FIG2_YEAST	111.00	109.96	12.34
SwissProt_40:TENS_CHICK	111.00	109.25	12.49
SwissProt_40:TRHY_HUMAN	111.00	108.49	12.64
SwissProt_40:YMH7_CAEEL	110.50	122.24	10.82
SwissProt_40:MANA_RHOMR	110.50	113.45	12.44
SwissProt_40:SEDI_YEAST	110.00	122.72	11.44
SwissProt_40:LEUK_HUMAN	110.00	121.23	11.71
SwissProt_40:CYP8_CAEEL	110.00	119.87	11.97
SwissProt_40:ITF2_MOUSE	110.00	116.64	12.60
SwissProt_40:CLPB_MCTU	110.00	114.54	13.02
SwissProt_40:NKCR_HUMAN	110.00	109.69	14.06
SwissProt_40:TEGU_HSVB	110.00	102.13	15.85
SwissProt_40:YK82_MCTU	109.50	115.42	13.67
SwissProt_40:PBPA_BACSU	109.50	113.31	14.14

seq\_name: SwissProt\_40:MUC2\_HUMAN

seq\_documentation\_block:

ID MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
AC Q02817; Q14878;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mucin 2 precursor (Intestinal mucin 2).  
GN MUC2 OR SMUC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Percent Similarity: 40.419 Percent Identity: 22.754

alignment\_block:

US-09-303-518d-463 x MUC2\_HUMAN

Align seg 1/1 to: MUC2\_HUMAN from: 1 to: 5179

```
269 ACATTGTCCTGTTTCCTGATCAGCGGCAAAATTCATTCCTCCCTTGAC 318
|||||
1520 ThrThrThrProSerProThrThrThrThrThrThrThrThrThrThr 1536
319 AACATGCTCCTCAGATTCGATTCGAGCAAGCCGGTAGTCCGTTGACGG 368
|||||
1536 rThrThrProSer...ProProThrThrThrThrProleThrProThrS 1552
369 ATTACGCTTTACCGCATCCATCGGAGGATAGCAACACATCCCGCG 418
|||||
1552 eThr...ThrThrLeuProProThrThrThrProSerProProPro 1566
419 ACGGTATGACGGGCGCACAGCGCGGCTATCCCGCTCCCAAGGCGG 468
|||||
1567 ThrThrThrThrThrThrThrThrThrThrThrThrProSerPro... 1580
469 AGGGATATATACAGCTAGCATAAAGCGCTTCCGCAAAATATCCGCT 518
|||||
1581 .....ProThrThrThrThrThrThrProSerProThrThrThrThr 1594
519 CAACCTGACGACACCGCCAGCGGACACGGCTTCCGACCGTTCC 568
|||||
1594 hrThrProProProThrThrThrThrProSerProProThrThrThr 1610
569 ACAATGCGCGGCTATGCTGACGAAGGAGTAGGCGGAGGATCAACGC 618
|||||
1611 ThrProPro.....ProThrThrThrThrProSe 1619
619 GCCACCGATACAGCCCGAGCTGGACAGATCGGGCAATCGCGCGAAGC 668
|||||
1619 rProProThrThrThrThrProleThrProProThrThrThrThrLeuP 1636
669 CTTCACGCGACTGACAGATATCGTCAAAACATCATCGCGCGCAGAG 718
|||||
1636 roProThrThrThrProSerProProProThrThrThrThrThrPro 1652
719 AAATTGTCGGCGAGCGATGCGGTGAGGTATAAGCAAGGCTCAAC 768
|||||
1653 ProThrThrThrProSerProThrThrThrThrProSerProThr 1669
769 ATTGCTGTCAGCAGCGCTTGGTCTGTTTCCACCGCAAAACAGATGGC 818
|||||
1669 eThrThrThrThrThr.....ProProProThrThrThrThrP 1681
819 GCGATCAACGATTGGCAGATATGGCGCAACTCAAGACTATCCGCGAG 868
|||||
1681 roSerProle.....ThrThrThrProSer 1690
869 CAGCA.....TC 876
1691 ProProThrThrThrMetThrThrProSerProThrThrThrProSe 1707
877 CGCGATTGGGAGTCCAAACCCCAATGCCGACAGCATAGAGCGGT 926
|||||
1707 rProleThrThrThrThrThrProSerSerThrThrThrProSerProp 1724
927 CAGCAATATCTTATGCGAGCCATCCCATCAAGAGGATTGGAGCTGCC 976
|||||
1724 roProThrThrMetThrThrProSerProThr..... 1734
977 GGGAAAAATACGGTTGGCGGCATCAGCGCACATCTGTCAAGCGGTGC 1026
1734 ..... 1734
1027 CAGATGGCGGATCGCATTCGCGAAAGGAAATCCGCGTCAGCGACAA 1076
|||||
```

```
1735 .....ThrThrProSerProProThrThrThrM 1744
1077 TTTTGGCGATGGCGCATACGCAATACCGTCCCTTACCAT 1120
1744 eThrThrLeuProProThrThrThrSerProleThrThrThrPro 1760
1121 .....CCGAAATATCCCTTCAACGTTGAGCAGCGTTACGGCAAGAA 1164
|||||
1761 LeuProProSerileThrProProThrPheSerProPheSerThr 1775
1165 AACATCAGCTCTCAACCGTGGCGCGCTCAACGGCAAAATGTCAACT 1214
|||||
1776 .ThrThrProThrThrProCys.....ValProLeuCysAsnT 1788
1215 GG 1216
1788 rp 1788
```

seq\_name: SwissProt\_40:FVB\_MOUSE

seq\_documentation\_block:

ID FVB\_MOUSE STANDARD; PRT; 819 AA.

```
AC O35601; Q922H3;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE FYN-binding protein (FYN-T-binding protein) (FVB-120/130) (p120/p130)
DE (SLP-76 associated phosphoprotein) (SLAP-130).
GN FVB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM FVB-120).
RC TISSUE=T-cell lymphoma;
RX MEDLINE=97352826; PubMed=9207119;
RA da Silva A.J., Li Z., de Vera C., Canto E., Findell P., Rudd C.E.;
RT "Cloning of a novel T-cell protein FVB that binds FYN and SH2-domain-
RT containing leukocyte protein 76 and modulates interleukin 2
RT production."
RL Proc. Natl. Acad. Sci. U.S.A. 94:7493-7498(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM FVB-130).
RC TISSUE=Hybridoma;
RX MEDLINE=99428514; PubMed=10497204;
RA Veale M., Raab M., Li Z., da Silva A.J., Kraeft S.-K., Werenowicz S.,
RA Morton C.C., Rudd C.E.;
RT "Novel isoform of lymphoid adaptor FYN-T-binding protein (FVB-130)
RT interacts with SLP-76 and up-regulates interleukin 2 production."
RL J. Biol. Chem. 274:28427-28435(1999).
CC -!- FUNCTION: ACTS AS A ADAPTER PROTEIN OF THE FYN AND SH2-DOMAIN-
CC CONTAINING LEUCOCYTE PROTEIN-76 (SLP76) SIGNALING CASCADES IN T
CC CELLS. MODULATES THE EXPRESSION OF INTERLEUKIN-2 (IL-2).
CC -!- SUBUNIT: INTERACTS WITH FYN AND SLP76.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: FVB-130 (SHOWN HERE) AND FVB-
CC 120; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEMATOPOIETIC TISSUES SUCH AS
CC MYELOID AND T CELLS, SPLEEN AND THYMUS. NOT EXPRESSED IN B CELLS,
CC NOR IN NON-LYMPHOID TISSUES. FVB-130 IS PREFERENTIALLY EXPRESSED
CC IN MATURE T-CELLS COMPARED TO FVB-120, WHEREAS THYMOCYTES SHOWED A
CC GREATER RELATIVE AMOUNT OF FVB-120.
CC -!- PWM: T-CELL RECEPTOR LIGATION LEADS TO INCREASED TYROSINE
CC PHOSPHORYLATION.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)





```

393 rOleuProAlaSerHisProAlaHisProProValProSerLeuProPro 409
1201 AAAAATGTCAAA...CTGGCAGACCAACGCCACCGG..... 1233
410 ArgAsnIleLysProProLeuAspLeuLysHisProIleAsnAspGluAs 426
1234 .AAGACAGCGCTACCGTTT...GACGGTAAAGGGTTTCGGAATTTGAGA 1279
426 nGlnAspGlyValMetHisSerAspGlyThrGlyAsnLeuGluGluG 443
1280 AGCAGCTGAATATGATACAGAGCTCGATATCAAGAATTATCGGGGGC 1329
443 InGluSerGluGlyGluThrTyfGluAspIleAspSerSerLysGluArg 459
1330 GGTATACCTAAGCGTAAAGCTGTTTTCATGTCGGAACCGAGATGGAGGT 1379
460 AspLysLysArgGluLys.....GluGluLysLysArgLeuGluLe 473
1380 TGATAGGAAGCTTAATAAATTGACAACCTCGTAGCAGGTGAGAAAAATG 1429
473 uGluArgLysGluGlnLys.....GluArgGluLysLysGluGlnGluL 488
1430 TTCAGGAACAGAGAGAGAGAGTACAGTAGTCAGTTTAAAGCCCATGCG 1479
488 euLysLysLysPheLysLeuThrGlyProIleGlnValIleHisHisAla 504
1480 CAACGAGAATGGGAAAAATAAAACAGGCTTAGATTTTAAATCATTTTATAGG 1529
505 LysAlaCysCysAspValLysGlyGlyLysAsnGluLeuSerPheLysGI 521
1530 TGGTGATATCANTAAGAAAGCACAGTAGTAACAGGAGCGCATAGTCAACC 1579
521 nGlyGlu..... 523
1580 GTGGTGATGACGGGTGATACACACACCTCGGCACCTGTATAACATGGG 1629
524 ....AspIleGluIleIleArgIleThrAsp..... 532
1630 GTTTATCAACGACAGTAGTGAATTAAGAACCTCGATGGAAGTTGGGAGGT 1679
533 .....AsnProGluGlyLysTrpLeuGI 540
1680 GAAACCAAAAAGCTGGGAAAGTAGTACCAACAGCACACCATGTTCCCAA 1729
540 YArgThrAlaArgLysSerTyfGlyTyfIleLysThrAlaValGluI 557
1730 AAGATTGGGATGAGGCTAGATTAAGGCTGAAGTTACTTCGGCTTGGAA 1779
557 leAspTyfAspSerLeuLysArgLysLysAsnSerLeuAsnAlaValPro 573
1780 AGTAGAATAATGCTTAAGGATAAT..... 1803
574 ProArgLeuValGluAspAspGlnAspValTyfAspValAlaGluGI 590
1804 .....AAATGCGAGGTACAGTAATTCGGGTATTAATAATAGAG 1843
590 nAspAlaProAsnSerHisGlyGlnSerGlySerGly.....G 603
1844 GATTTACCGAACCTAATAAGACAGCATATCCCATTTATGAA 1884
603 lyMetPheProProProThrAspAspGluIleTyfAsp 616
seq_name: SwissProt_40:DAN4_YEAST

```

seq\_documentation\_block:

```

ID DAN4_YEAST
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
GN DAN4 OR YJR151C OR J2223.

```

```

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RA Scarce T.; Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=21113168; PubMed=11160904;
RA Cohen B.D., Sertil O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites.";
RL Nucleic Acids Res. 29:799-808(2001).
CC -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SRP1 / TIPI FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z49651; CA89684.1; .
DR SGD; S0003912; DAN4.
DR InterPro; IPR000992; SRP1_TIP1.
DR Pfam; PF00660; SRP1_TIP1; 1.
DR PROSITE; PS00724; SRP1_TIP1; 1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;

```

alignment\_scores:

```

Quality: 153.50 Length: 429
Ratio: 0.717 Gaps: 12
Percent Similarity: 49.883 Percent Identity: 22.611

```

alignment\_block:

US-09-303-518d-463 x DAN4\_YEAST ..

Align seg 1/1 to: DAN4\_YEAST from: 1 to: 1161

```

77 ATTGGCAACGATCCCTTTATCCGGCAGGTCTCGACCGTCAGCATTC 126
||||| :||||| ||| :||||| |||
117 lleyrThrAlaIleProThrSerThrSerThrThrThrLysSerSe 133
127 GAACCCGACGGAAATACCACTATTCGGCAGCAGGGGGAGCTGCCNA 176
|::: ||| ||||| :||| |||
133 rThrSerThrThrProThrThrThrIleThrSerThrThrThrThrS 150
177 GCGCAACGCCATATCGGATTTGGGAACATACAAAGCCATCGATTGGGCC 226
||||| :||| :||| :||| :|||
150 erThrThrProThrThrSerThrThrThrThrThrProThrThrSerThr 166
227 ACCTGATGATTCACAGCGCGCGCTGAAGAAATATCGGCTACATGTC 276
||| :||| :||| :||| :|||
167 Thr.....SerThrThrProThrThrSerThrThrSe 177
277 CGCTTTCCGATCAGCGGCACAAATTCATTCGCCCTTCGACACACCATGC 326
|::: ||| |||:||||| ||| :|||:|
177 rThrThrProThrThrSerThrThrThrThrThrProThrThrSer.... 192

```



KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 FT SIGNAL: Multigene family.  
 FT CHAIN 1 21 POTENTIAL.  
 FT DOMAIN 22 1367 GLUCOAMYLASE S1/S2.  
 FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

## alignment\_scores:

Quality: 146.00 Length: 438  
 Ratio: 0.737 Gaps: 16.  
 Percent Similarity: 45.205 Percent Identity: 23.288

## alignment\_block:

US-09-303-518d-463 x AMYH\_YEAST ..

Align seg 1/1 to: AMYH\_YEAST from: 1 to: 1367

```

20 TATCCCTTATCTGCTCCATCTAGTGGCAGTGTGCCTGC.....CGATGCAT 63
192 PheProGlyPhe.....TyrTrpAsnIleAspCysAspAsnAsnCysG1 206
64 GCACACGGCTCAGATTGGCAACAGATCCCTTTATCCGGCAGGTTCTCGA 113
206 yGlyThr.....LysSerSert 212
114 CGGTGAGATTTCGACCGGAGGGAATACACATATTCGGCAGCAGG 163
212 hrThrThrSerThrSerGluSerThrThrThrThrThrThrThrThr 228
164 GGGAGCTTCCGACGCGCAACGGCCATATCGGATTGGGAACATACAAAGC 213
229 GluSerThrThrThrThrThrThrThr..... 238
214 CATCATGTTGGGCCACCTGATGATTCAACAGCGCGCGGTGAAGGAATAT 263
239 .....GluSers 241
264 CGGCTACATTGCGCTTTTCGGATCAGCGGCACAAATTCATCGCCCT 313
241 erThrThrThrThrSer.....ThrSerGluSerThrThrSer 254
314 TCACAAACATCCCTCACATCCGATTCGTACG.....AA 348
255 SerThrThrAlaProAlaThrProThrThrThrThrThrThrThrThr 271
349 GCGGTAGTCCGTTGACGGATTCAGCCTTTACCGCATCCATTCGGACGG 398
271 sProThrProThrThrThrThrSerCys.....ThrL 282
399 ATACGACACCATCCGCGG..... 418
282 ysGluLysProThrProThrHisAspThrThrProCysThrLysLys 298
419 .....ACGGCTATGACGGCCACAGGGCGGCGGTATCCCGCTCCC 459
299 LysThrThrThrSerLysThrLysThrLysThrThrThrThrThrThr 315
460 AAAGCGCGGAGGATATACAGCTACGACATAAAAGGCGTTGCCCAAAA 509
315 oThrProSerSerThrThrGluSerSerSerAlaProValThrP 392
510 TATCGCCCTACCTGACCGCAACCGGACCGACCGACACGGCTTGGCG 559
332 roSerSerThrThrGluSerSerSerAlaProValThrSerThr 348
560 ACCGTTTCCACAAATCCCGCGGTATGCTCACCAAGGAGTAGCGCACGA 609
349 ThrGluSerSerAlaProValProThrProThrProSerSerThrThr 365
610 TTCAAACGGCGCCACCGATACAGCCCGAGCTGGACAGATCGGCAATGC 659

```

```

365 userSerSerAlaProValThrSer...SerThrThrGluSerSerSera 381
660 CGCCGAAGCCTTCAACGGCCTACTGCAGATATCGTCA.....AAA 697
381 laProValThrSerThrThrGluSerSerSerAlaProValProThr 397
698 ACATCATCGCGCGGAGGAGAAATTGTCGGCGCAGCGATGCGGTGCAG 747
398 ProSerSerThrThrGluSerSerSerAlaProValThrSerSerTh 414
748 GGTATAAGCGAAGGCTCAAAACATTGTGCTCATCGACGGCTTGGGTCT 797
414 rThrGluSerSerAlaProValThrSerSerThrThrGluSerSers 431
798 TTCACCGGAAAACAGATGCGCGGCATCAACGATTGGCAGATATGCGCG 847
431 erAlaProValThr.....SerSerThr.....Thr 439
848 AACTCAAGACTATGCGCAGCAGCATCGGATTTGGCGAGTCCCAAAAC 897
440 GluSerSerSerAlaProValThrSerSerThrThrGluSerSerSer 456
898 CCCAATGCCGCACACAGCATAGAACCGTCAGCAATA.....TCCT 938
456 aProValPro.....ThrProSerSerThrThrGluSers 469
939 TATGGCAGCCATCCCATCAAGGGATTGGAGCTGTCGGGGAAATACG 988
469 erSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 485
989 GCTTGGCGGCATCACGGCACATCTCTCAAGCGGTGCGCAGATGGCGCG 1038
486 ProThrProSerSerThrThrGluSerSerSerAlaProValThrSe 502
1039 ATCGCATTTGCCGAAAGGAAATCCGCG.....TCAG 1070
502 rSerThrThrGluSerSerAlaProValProThrProSerSerSert 519
1071 CGACAATTTGCCGATCGGCATACGCCAAATACCCGTCCTCCCTTACC 1120
519 hrThrGluSerSerSerAlaProAlaProThrProSerSerThrThr 535
1121 CCGGAAATATCCGTTCAAACTTGAGCAGCGCTTACGGCAAGAAACATC 1170
536 GluSerSerSerAlaProValThrSerSerThrThrGluSerSerSer 552
1171 ACCT...CCTCAACCGTGGCGCGCTCAACGGCAAAATGTCAACTGCG 1217
552 aProValProThrProSerSerSerThrThrGluSerSerThrPro 569
1218 AGACCAACGCCACC 1231
569 alThrSerSerThr 573
seq_name: SwissProt_40:NKCR_MOUSE
seq_documentation_block:
ID NKCR_MOUSE STANDARD; PRT: 1453 AA.
AC P30415;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE related protein) (NK-TR protein).
GN NKTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93133824; PubMed=8421688;

```









DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;  
KW Nuclear protein.  
FT DOMAIN 51 56 NUCLEAR LOCALIZATION SIGNAL  
FT DOMAIN 92 146 (POTENTIAL).  
FT DOMAIN 386 504 HIS-RICH.  
FT DOMAIN 558 851 SER-RICH.  
FT NP\_BIND 564 572 PROTEIN\_KINASE.  
FT BINDING 587 587 ATP (BY SIMILARITY).  
FT ACT\_SITE 685 685 ATP (BY SIMILARITY).  
FT CONFLICT 69 69 BY SIMILARITY.  
FT CONFLICT 223 223 D -> E (IN REF. 1).  
FT CONFLICT 283 283 V -> A (IN REF. 1).  
FT CONFLICT 492 492 I -> S (IN REF. 1).  
FT CONFLICT 584 584 V -> A (IN REF. 1).  
FT CONFLICT 603 603 V -> E (IN REF. 3).  
FT CONFLICT 612 612 L -> R (IN REF. 3).  
FT CONFLICT 670 670 G -> GG (IN REF. 3).  
FT CONFLICT 680 680 MISSING (IN REF. 3).  
FT CONFLICT 717 717 MISSING (IN REF. 3).  
SQ SEQUENCE 952 AA; 108837 MW; 9CBDE8D459D0713D CRC64;

alignment\_scores:  
Quality: 141.00 Length: 473  
Ratio: 0.662 Gaps: 25  
Percent Similarity: 45.032 Percent Identity: 21.353

alignment\_block:  
US-09-303-518d-463 x KP58\_DROME ..  
Align seg 1/1 to: KP58\_DROME from: 1 to: 952

393 GGACGGATACGACACCA.....TCCGCCGACGGCT 424  
|||||.....  
10 GlyGlnLeuArgSerProAspValHisTyrHisSerArg..... 23:  
425 ATGACGGCCACAGCGCGCGCTATCCGCTCCCAAGGCGCGAGGGA. 473  
.....GlyGluGluAspLuhisGluGlyA 32  
474 .....TATATACGTACGACATAAAGCGTTGCCCAAAA 509  
|||||.....  
32 spAlaAspAlaLeuTyrIleGln..... 39  
510 TATCGCGCTCAACCTGACCGACACCGCAGCACCAGCGGCTTGCCG 559  
|||||.....  
40 ...ProGlnAlaSerArgLysSerGly..... 50  
560 ACCGTTTCCCAATCCGCGCGCTATGTGACGCAAGGAGTAGCGACGGA 609  
|||||.....  
51 .....ProArgArgGluLysLysHisSerArgGluArgArgArg 65  
610 TTCAACCGCCACCCGATACAGCCCGCGCTGACAGATCGGCGCAATGC 659  
|||||.....AspaspValGlyGlyAla 74  
65 isLysGluArg.....CGTCAAAA 697.  
660 CGCGAAGCCCTTCAACGCGCTGCAGATAT.....CGTCAAAA 697.  
75 AlaLeuAlaLeuGluArgAspHisArgTyrAspTyrArgSerArgGluG 91  
698 ACATCATCGCGCGCGAGGAAATTTGCGCGCAGCGCGATGCGGTGCAG 747  
|||||.....  
91 uHisTyrHisHisGlnArgGlu...ArgSerSerAsnAlaAlaAla 107  
748 GGTATAACGGAAGGCTCAACATTCCTGTCTATGACCGCTTGGCTGCT 797

107 laTyr.....AlaLysHis.....HisLeuGlyHisAla 116  
798 TTCACCGAAACAAAGATGGCGGCATCAACGATTTGGCAGATATGGCGC 847  
|||||.....  
117 TyrHisTyrProGlnProGlnGlnGln..... 127  
848 AACTCAAGACTATGCCGACGACCCCATCGCGATTGGCGAGTCCCAAAAC 897  
|||||.....ProProAlap 134  
128 ....GlnProLeu.....  
898 CCCAATGCCGACCAAGCATAGAGCGTCAGCAATATCTTTATGGCAGC 947  
|||||.....  
134 roSerTyrAlaAlaHisHisTyrHisHisGlnHisLeuSerGlyAla 150  
948 CATCCCATCAAGGGATTGG...AGCTGTCCGGGAAATACGGCTTGG 994  
|||.....  
151 ArgAlaAlaProArgGluTyrHisSerTyrProSerGlyTyrHisSerG 167  
995 GCGGCATCAGCGGCACATCTGTCAAGCGGTCCGAGATGGCGCGCATCGCA 1044  
|||.....  
167 ySerArgHisGly.....AspTyrP 174  
1045 TTGCGGAAAGGAAATCCGCGTCAGCGACAATTTGCGCATCGCGCAT 1094  
|||||.....  
174 roMetGluGluProThrArgArg.SerSerLysTyrAlaGluSerLysas 190  
1095 CGGCAAA.....  
190 palagluSerLeuGluGlnAspLeuArgSerArgLeuLeuLysLysArgH 207  
1102 .....TACCGTCCCTTACCATTC...CGAAATATCGTTCAAACTTG 1143  
|||||.....  
207 isAsnTyrValLysAspTyrGluThrGluGluAsnTyrGluHisArgVal 223  
1144 GAG.....CAGCGTTACGGCAAGAAACATCACCTCCCTCAACCGT 1184  
|||||.....  
224 GluArgSerAspArgArgGluGlyGlyArgLysGluArgGluArgThrVa 240  
1185 GCGCGCGTCACACGGCAAAATGCAAACTGCGAGACCAACGCCACCGCA 1234  
|||||.....ArgHisAspA 251  
240 lArgSerThrHisLysGlnAsn.....  
1235 AGACAGCGGTACCGTTTACGGTAAAGGTTTCCGAATTTGAGAGACAC 1284  
|||||.....  
251 rgValleGluLeuLeuAspSer.....ProGluGlnGluHisHis 264  
1285 GTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGCGGTAT 1334  
|||||.....  
265 HisGlnHisGlnHisLys..... 270  
1335 ACCTAAGGCTAACCTGTGTTCATCGCAACCGAGATGG..... 1374  
|||.....SerHisArgSerLysTyrArgGluGlu 280  
1375 .....GAGGTTGATAGGAAGCTTAATAATTGACAACCT 1407  
|||||.....  
280 algluValleArgArgLysValProGluAspLeuLeuLeuAlaArg 296  
1408 CGTGACAGGTGGGAAAAATGTTTCAGAAACGAGAAAGAGGTACAG 1457  
|||||.....  
297 ArgGluLysLeuLeuAlaAlaGluArgGluSerArgGlnArgLysGlnPh 313  
1458 TAGTCAGTTTAAAGCCCATCGCAACGAGAAATGG.....G 1492  
|||||.....  
313 rAlaArgGluGluLeuGluAlaArgGluLeuLeuArgGluArgAsnG 330  
1493 AAAATAAAACAGGTTTAGATTATTAATCATTTTATAGGTGCTCATATCAAT 1542  
|||||.....  
330 luHisSerAspAlaLeu..... 335  
1543 AAGAAAGGCACAGTAAACAGGAGGCATGCTAACCCGCTGTGTATGACG 1592  
|||.....



```

1087 GCGCATACGCAAAATACCGTCCCTTACCATTCCTCCGAAATATCCGTTCC 1136
    :::      :::      |||      |||||||:::      |||::
315 SerGluArgArgGluThrProProArgProHisSerArgLysArgArgAs 331
1137 AAATTCGACGAGCGTTACGCGCAAGAAACATCACCTCTCAACCGTGC 1186
    :::      |||||||      |||||||      |||
331 pThr..GlyAla.....HisHisArgHisTrpArgAr 341
1187 CGCGTCAACGCGCAAAATGTCAAACTGGCAGA.....CCAACGCCAC 1230
    :::|||:::      :::      |||      :::
341 gArgArgArgValArgHisArgGluGlyAlaLeuProAlaAlaHisP 358
1231 CCGAAGACAGCGGT 1244
    ||:|||||
358 roAspAspArgArg 362

```

seq\_name: SwissProt\_40:DAB\_DROME

seq\_documentation\_block:

```

ID DAB_DROME STANDARD; PRT: 2411 AA.
AC P98081;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Disabled protein.
GN DAB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194063; PubMed=7680635;
RA Gertler F.B., Hill K.K., Clark M.J., Hoffmann F.M.;
RT "Dosage-sensitive modifiers of Drosophila abl tyrosine kinase
RT function: prospero, a regulator of axonal outgrowth, and disabled, a
RT novel tyrosine kinase substrate.";
RL Genes Dev. 7:441-453(1993).
CC -!- FUNCTION: TOGETHER WITH ABL INVOLVED IN EMBRYONIC NEURAL
CC DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: CNS AXONS AND BODY WALL MUSCLES.
CC -!- DEVELOPMENTAL STAGE: EMBRYONIC AXONOGENESIS.
CC -!- PTM: PROBABLY PHOSPHORYLATED BY THE ABL TYROSINE KINASE.
CC -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L08845; AAB08527.1; ALT_SEQ.
CC PIR: A46299; A46299.
CC FlyBase: FBgn0000414; Dab.
CC InterPro: IPR000050; PID_domain.
CC Pfam: PF00640; PID; 1.
CC SMART: SM00462; PTB; 1.
CC PROSITE: PS01179; PID; 1.
CC Alternative splicing; Phosphorylation; Repeat.
CC DOMAIN 46 196
CC REPEAT 1689 1801
CC REPEAT 1689 1700 ALTERNATE ARG AND ACIDIC RESIDUE.
CC REPEAT 1740 1750 ALTERNATE ARG AND ACIDIC RESIDUE.
CC REPEAT 1791 1801 ALTERNATE ARG AND ACIDIC RESIDUE.
CC MOD_RES 111 111 PHOSPHORYLATION (BY ABL) (PROBABLE).
CC MOD_RES 482 482 PHOSPHORYLATION (BY ABL) (PROBABLE).

```

```

FT MOD_RES 1662 1662 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1667 1667 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1701 1701 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1704 1704 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1713 1713 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1739 1739 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1826 1826 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT MOD_RES 1961 1961 PHOSPHORYLATION (BY ABL) (PROBABLE).
FT VARSPIC 462 673 MISSING (IN SHORT ISOFORM)
SQ SEQUENCE 2411 AA; 264046 MW; 579AB9C0243D5FD6 CRC64;

```

alignment\_scores:

```

Quality: 138.50 Length: 558
Ratio: 0.523 Gaps: 38
Percent Similarity: 47.491 Percent Identity: 23.835

```

alignment\_block:

US-09-303-518D-463 x DAB\_DROME ..

Align seg 1/1 to: DAB\_DROME from: 1 to: 2411

```

266 GCTACATTCGCGCTTTCCGATCAGGGCACAATTCATTCCG...CC 312
||||| ||||| ||| |||||::: |||
1405 AlaThrSerSerAlaAlaProSerArgGlyAlaProGlyLeuHisThrPr 1421
313 TTCG.....ACAACCATGCCTCATTCCGATTCTGACGA 347
||||| ::||| ||| ||||| |||
1421 oSerGlnPheAsnAspValSerThrSerPro...ilePro...LeuGlnL 1436
348 AGCCG.....GTAGTCCCGTTGACGGATTACCCCTTTACC 382
||||| ||||| ||||| ||||| |||
1436 ysProGlyMetGlyProSerProValProSerGlnLeuSerAlaValSer 1452
383 GCATCCATTGGGACGATACGAACACCATCCGCGCAGCGTATGACGG 432
||||| ||| ||| ||||| |||
1453 GlnLeuIleAspThrAlaThrLys.....GlnMetMetG 1464
433 CCACAGGCGG.....GCGGCTATCCCGCTCCCAAGG 464
|||::: |||::: |||::: |||::: |||
1464 yAspLysAspArgGluLysGlnSerTrpAlaThrPheAspSerProLysA 1481
465 CGCAGGAGGATATACAGCTACGACATAAAAGCGTTGCCCAAAATATCC 514
|||::: ||| ||| ||| |||
1481 laLysGlyLysAlaArgLeuThrLeuProProProProProAlaSer 1497
515 GCCTCAACCTCAGCGCAACCGCAGCCGACACCGCTTGCCTG... 559
::: ||| ||| ||| ||| |||
1498 AsnThrSerGlnProAspThr.....GluAsnArgLeuAlaValar 1511
560 .....ACCGTTTCCACAATCCGCGCGCTATGCTGACGCAAGG 596
|||::: |||::: |||::: |||::: |||
1511 gileProGlyMetThrAlaGlyGlnSerAsnSerValValGlyArgArgA 1528
597 AGTAGGACGAGGATTAACAGCGCCACCGATACAGCCCGAGCTGGACA 646
:: ||||| |||::: ||| ||| |||
1528 rgSerAlaThrThrSerSerSerSerArgAspLeuSerPro...TrpAsp 1543
647 GATCGGCAATGCGCGCAAGCCCTTCAACGCGCTGCAGATATCGTCAAA 696
|||::: |||::: |||::: |||::: |||
1544 AspGluThr.....ProGluTyrLeuLysArgArgGlnLeuAlaAla.. 1557
697 AACATCATCGCGCGCAGGAGGAAATTTGTCGCGCAGCGGATGCGCG... 742
::: |||::: |||::: |||::: |||
1558 .AlaGlnMetAlaHisProHisGlnProMetGlnAlaProProGlnH 1574
743 .....TGCAAGGTATTAAGCGAAGGCTCAACA 769
|||::: |||::: |||::: |||::: |||
1574 isThrAspArgHisGlyTyrTyrMetArg.HisAlaArgArgMetAsnSe 1590
770 TTGCTGTGCAT.....GCACGCGCTTGGGT 792
|||::: |||::: |||::: |||::: |||
1590 rCysAspGluAspTyrAspTyrAspGlyGluPheValAlaArgAspG 1607

```

```
793 CTGCTTTCCACCAAAACAAAGATGGCGCCATCAACGATTTGCGACATAT 842
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1607 lnProGlnHisGlnGlnGln.....GlnArgLysPheLysHis 1619
843 GCGCCTCACTCAAGACTATGCGCGAGC..... 869
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1620 Gly.....LeuSerArgSerArgAspAsnPheGluLeuGluSe 1632
870 .....AGCATCCG.....CGATTGGCAGCTCCAAAA..... 896
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1632 rProSerTrpTyrHisHisProAlaHisHisThrTrpSerProGlnGln 1649
897 ..CCCCAATCCCGACAGCAGATAGAGCGCTCAGCAATAT..... 935
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1649 leGluGlnAlaArgValArgSerPheAspArgThrAlaTyrGluArgSer 1665
936 CTTTATGCGACCCATCCCATCAAGAGGATTCGAGCTGTCGCGGGGAAAT 985
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1666 SerTyrGlyProProProProlleTyrAspLysArg.....GlyGlnLe 1680
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
986 ACGGCTTGGCGGCATCACCGCATCTCTCTCAAGCGGTGCGAGATGGGC 1035
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1680 uArg...GlyLysTyrArgGly.....AspHisA 1689
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1036 CGATCGCATTCGCGAAGGAAATCCGCGTCAGCGACAAATTTGCGGA 1085
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1689 rgAspArg.....GluArgGluArgAspArgGlu.TyrArgAs 1703
1086 TGCGGCATACGCCAAATACCGCTTACCATTCGCCAAATATCGGTT 1135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1703 pTyrAlaArgProSerTyrAspPheAspTyr.....GluAsnVal.... 1716
1136 CAAACTTGGACGCGTTACGGCAAGAAACATCCTCTCTCAACCGTG 1185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1717 .....TyrGluGluArgGlyArgSerProLeu.....AlaTyr 1728
1186 CCGCGCTCAACGCCAAATGTCAACTGCGAGAC..... 1221
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1729 LysProGlyArgGlyGlyArgPheTyrLeuTyrAspArgGluArgAsp 1745
1222 .....CAACGCCACCCGAGACAGCGGTACCGTTTGACGGTAAAGGT 1264
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1745 gAspArgGluArgAspProLysSer.....PheAspArgGluSerL 1759
1265 TTCGGAATTTGAGAACGACCTGAAATATATATACGAACCTCATATCAA 1314
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1759 euGluSerTyrGluSerAlaThrArg..... 1768
1315 GAATTATCGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAA 1364
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1769 ArgArgSerPheGlySerGlyAsnAspValTyrGlySerLeuAspSer 1785
1365 ACCGAGATGGGAGTTGATAGAACCTTAATAATGACAACTCGTGAGC 1414
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1785 gAspAspTyrArgGlyAspArgGluArgAspArgGluArgAspArgGlu 1802
1415 AGGTGGAGAAAATGTTCCAGGAACGACAGAGAGTACAGTAGT... 1461
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1802 lnMetLys.....ThrArgSerLeuArgLysProThrThrThrSerGly 1816
1462 .....CAGTTTAAAGCCCATGCGCAACGAGA 1487
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1817 LysLeuArgIleSerGlyAspIleAspTyrGluGlnAspSerGluGlnAs 1833
1488 ATGGGNAATAAACAGG.....TTAGATTTTAATCATCTTTATAG 1528
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1833 pPheGlnGlnArgSerGlyValArgSerLeuGlnLeuProAsnGlnLeuG 1850
1529 GTGGTGAT...ATCAATAAGAGGACAGTAAACAGGAGGCGCATAGTCTA 1575
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1850 lyGlyAspValValLeuProSerAsnAlaValAlaGlyProGlnArgPhe 1866
```

```
1576 ACCGCTGTGTGATGACGGGTGATACAAACACCTCGGCACCTGATAACA 1625
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1867 Alalys.....SerSerGlySerValProGlyArgAr 1877
1626 TGGG.....GTTTATCAAGCGACAGAGTGGAATTA 1654
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1877 gGlyThrProCysLeuValArgAsnLeuGlySerValGlnProAlaLeuL 1894
1655 AAAAGCCTGTAGGAAGTTGG 1674
1894 euLysProLysGlyAspTrp 1900
seq_name: SwissProt_40:PODX_RABIT
seq_documentation_block:
ID: PODX_RABIT STANDARD; PRT; 551 AA.
AC Q28645; 2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Podocalyxin-like protein 1 precursor.
GN PODXL OR PCLPL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RC STRAIN=NEW ZEALAND WHITE;
RX MEDLINE=96094343; PubMed=7493982;
RA Kershaw D.B., Thomas P.E., Wharram B.L., Goyal M., Wiggins J.E.,
RA Whiteside C.I., Wiggins R.C.;
RT "Molecular cloning, expression, and characterization of podocalyxin-
RT like protein 1 from rabbit as a transmembrane protein of glomerular
RT podocytes and vascular endothelium."
RL J. Biol. Chem. 270:29439-29446(1995).
CC -!- FUNCTION: Functions as an antiadhesin that maintains an open
CC filtration pathway between neighboring foot processes in the
CC podocyte by charge repulsion.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: Glomerular epithelium cell (podocyte).
CC -!- PTM: Sialoglycoprotein.
CC -!- SIMILARITY: BELONGS TO THE PODOCALYXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U35239; AAC48489.1; -.
CC Glycoprotein; Signal; Transmembrane.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 551 PODOCALYXIN-LIKE PROTEIN 1.
CC DOMAIN 22 452 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 453 473 POTENTIAL.
CC DOMAIN 474 551 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 145 145 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 551 AA; 57040 MW; E9B8AE168CDFB8C5 CRC64;
```

alignment\_scores:  
Quality: 138.00 Length: 384  
Ratio: 0.754 Gaps: 23  
Percent Similarity: 47.656 Percent Identity: 26.302

alignment\_block:  
US-09-303-518D-463 x PODX\_RABIT ..

Align seg 1/1 to: PODX\_RABIT from: 1 to: 551

```
122 ATTTCGAACCCGACGGGAAATACACCTATTTCGCGACGAGGGGAGCTT 171
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
15 LeuSerProProSerLeuSerGlnGluLysSerProGlnProGlyProTh 31
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
172 GCNAGCCGCAACGGCCATATCGGATTTGGGAAACATACAAAGCCATCAGT 221
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
31 rProMetaLathrSerThrSer.....ThrArgProAlaProAlas 45
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
222 GGGCCACCTGATGATTCAACAGGGCGCGGTGAAGGAAATATCGGCTACA 271
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
45 eAlaPro.....AlaProLysSerSerValAlaAlaSer 56
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
272 TTGTCGCGCTTTTCGATCACGGGCACAAATTCATTCGCCCTTCGACAA 321
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
57 ValProAla.....GluGlnAsnThrTh 64
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
322 CATGCTCACATTCGGATTTCTGAGGAGCGCGTAGTCCCGTTGACGGATT 371
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
64 r.....PrometThrThrLys.....AlaProAlaThrGlns 75
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
372 CAGCCTTTACGCATCCATTCGGAGGATACG...ACACATCCCGCG 418
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
75 erPro...SerAlaSerProGlySerSerValGluAsnSerAlaProAla 90
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
419 ACGGCTATGACGGGCACAGGGCGGCGCTATCCCGCTCCCAAGGCGCG 468
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
91 GlnGlySerThrThrThrGlnGlnSerLeuSerValThrThrLysAlaG 107
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
469 A.....GGGATATATACAGT...ACGACATAAAGGCGT 500
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
107 uAlaLysAspAlaGlyGlyValProThrAlaHisValThrGlySerAla 124
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
501 TGCCC.....AAATATCCGCTCAACCTGACCGACCAACC 535
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
124 rgProValThrSerGlySerGlnValAlaAlaGlnAspProAlaAlaSer 140
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
536 GCAGCAGCGGACAGCGGCTGCCACCGCTTCCACAAATCGCG..... 577
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
141 LysAlaProSerAsn.....HisSerIleThrThrLysProLeuAlaTh 155
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
578 ...GCGTATGCTGACCAAGGAGTAGGCGGACGATTCAAACCGGCAC 623
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
155 rGluAlaThrSerGlnAlaProAlaGlnThrThrAspValGlyThrProG 172
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
624 CCGATACAGCCCGAGCTGGACAGATCGGGCAATGCGG..... 661
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
172 lyProThrAlaProProValThrAsnSerThrSerProAspLeuLeuGly 188
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
662 .....CGAAGCCTTCAACGGCACTGCAGATATCGTCAAAACATC 702
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
189 HisAlaThrProLysProSerGluGlyProGlnLeuSerPheProThrAl 205
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
703 ATC..GGCGGCGCAGAGAAATGTGCGGCGCAGCGCATCGCGTGAGGGTA 751
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
205 aAlaGlySerLeuGlyProValThrGlySerGlyThr..... 217
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
752 TAAGCGAAGGCTCAACACATCTCTCATGCGCGCTGGTCTGCTTTC 801
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
218 .....GlySerGlyThrLeuSer 223
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
802 ACC...GAAACAAGATGGCGCATCAACGATTTGGCAGATATGGCGCA 848
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
224 ThrProGlnGlyLysProAlaThrLeuThrProValAlaSerSerAlaG 240
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
849 ACTCAAGACTATG.....CCGAGCAGCAGCATCCGCGATTGGCAGTCC 892
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
240 u..ThrGlnGlyMetProSerPrometProProSerProAlaSerProSer 256
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
893 AAACCCCCAATGCGGCACAGGCATAGACCGCTCAGCAATATCTTATG 942
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
257 SerSerProPheProSer..... 262
```

```
943 GCACCATCCCATCAAGGATTGAGCTGTCGCGGGAATAATACGGCTT 992
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
263 .SerProSerProSerProAlaLeuGlnProSerGlyProSerAlaAlaG 279
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
993 GGCGGCGCATCACGGGCACATCTGTCAAGGGTGCAGATGGGCGCGATCG 1042
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
279 lyThrGluasp.....ThrThrGlyArg..... 286
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
1043 CATTCGCGAAGGAAATCCGCGTCAAGGACAAATTTTCCCGATGCGG... 1090
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
287 .....GlyProThrSerSerSerThrGluLeuAlaSerThrAl 299
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
1091 .....CATACGCCAAATACCGCTCCCTTACCATTCGCCGAATATCGGTC 1136
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
299 aLeuHisGlyProSerThrLeuSerProThr..... 309
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
1137 AAACCTGGAGCAGCTTACGGCAAGAAACATCA.....CCTCCTCAA 1180
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
310 .....SerAlaValArgAspGlnArgValSerCysGlyProGlu 323
    :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
seq_name: SwissProt_40:MAP4_MOUSE
seq_documentation_block:
ID MAP4_MOUSE STANDARD; PRT; 1125 AA.
AC P27546;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 4 (MAP 4).
GN MAP4 OR MTAP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042100; PubMed=1718985;
RA West R.R., Tenbarg K.M., Olmsted J.B.;
RT "A model for microtubule-associated protein 4 structure. Domains
defined by comparisons of human, mouse, and bovine sequences.";
RL J. Biol. Chem. 266:21886-21896(1991).
CC -!- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES
MICROTUBULE ASSEMBLY.
CC -!- TISSUE SPECIFICITY: TESTIS, STRIATED AND CARDIAC MUSCLE
CC -!- PTM: PHOSPHORYLATION OF THE PRO-RICH REGION IN THE C-TERMINUS
NEGATIVELY REGULATES MAP-4 ACTIVITY TO PROMOTE MICROTUBULE
ASSEMBLY.
CC -!- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M72414; AAA16372.1; -.
DR PIR; B41206; B41206.
DR MGD; MGI:97178; Mtap4.
DR InterPro; IPR001084; Tubulin-bind.
DR Pfam; PF00418; tubulin-binding; 4.
DR PROSITE; PS00229; TAU_MAP; 3.
KW Microtubules; Repeat; Phosphorylation.
FT REPEAT 896 926 TAU/MAP MOTIF.
FT REPEAT 965 995 TAU/MAP MOTIF.
FT REPEAT 996 1026 TAU/MAP MOTIF.
SQ SEQUENCE 1125 AA; 117675 MW; 73047432A329AA1D CRC64;
```

alignment\_scores:

Quality: 138.00

Length: 632

Ratio: 0.502 Gaps: 22  
Percent Similarity: 43.513 Percent Identity: 20.411

## alignment\_block:

US-09-303-518D-463 x MAP4\_MOUSE ..

Align seg 1/1 to: MAP4\_MOUSE from: 1 to: 1125

```
90  TCCTTTATCCGCGAGGTCTCCACCGTCAGCATTTTCGAACCCGCGGGA 139
|||||  |||||:  :::::  |||||:
543 SerMetAlaProAlaSerAspLeuAlaLeuProLeuGluThrLysValAl 559
|||||  |||||:  :::::  |||||:
140 AATACACCTATTCCGACAGCGGGGAGCTTGCNAGCGCAACGCCAT 189
|||||  |||||:  :::::  |||||:
559 aThrValProLleLysAspLysGlyThrValGlnThrGluGluLysProA 576
|||||  |||||:  :::::  |||||:
190 ATCGGATTGGG.....AAACATACAAAGCCATCAGTTGGG 224
:::  :::  |||||  |||||:
576 rGluAspSerGlnLeuAlaSerMetGlnHisLysGlyGlnSerThrVal 592
225 CCACCTGATGATTCAACAGG..CGGCGGTTGAAGGAAATATCGGCTACATT 273
|||||  :::::  |||||:  |||||:
593 ProProCysThrAlaSerProGluProValLys...AlaAlaGluGlnMe 608
274 GTCGCTTTTCGATACGCGGCACAAATTCATT..... 307
:||||:  |||||  :::  :::
608 tSerThrLeuProIleAspAlaProSerProLeuGluAsnLeuGluGlnL 625
308 ..CGCCCTCGACACACCATCGCTCACATCCGATTCTG..... 343
:::  :::::  |||||  |||
625 ySgluThrProGlySerGlnProSerGluProCysSerGlyValSerArg 641
344 .....ACGAAGCGGTAGTCCGCTGACGGATTACGCTTTACCGC 384
::|||  ::|||  |||||  ::|||
642 GlnGluAlaLysAlaValGlyValThrGlyAsnAspIleThr... 657
385 ATCCATTGGGACGATACGAACACCATCCGCGGACGCTATGACGGGCC 434
|||  |||  |||||
658 .....ThrProAsnLysGluProPro..... 665
435 ACAGGCGCGGCTATCCGCTCCCAAGGCGCGAGGATATACAGCT 484
|||  |||||  :::::  |||||
666 .....ProSerProGluLysLysAlaLysProLeuAla 676
485 ACGACATAAAGCGGTGCCCCAAATATCCGCTCAACGTCACCGACAC 534
|||||  |||  |||  ::|||
677 ThrThrGlnProAlaLysThrSerThrSerLysAlaLysThrGlnProTh 693
535 CGGACACCGGACAGCGGCTTCGCGACCGCTTTC..... 568
|||  |||  :::  |||||:  |||||
693 rSerLeuProLysGlnProAlaProThrThrSerGlyGlyLeuAsnLysL 710
569 .....ACAAATCCGCGCCTATGCTGACGCAAG 595
710 yProMetSerLeuAlaSerGlySerValProAlaAlaProHisLysArg 726
596 GAGTAGGCGACGATTCAAACCGCCACCGCATACAGCCCGCGCTGGAC 645
|||||  :::::  |||||  |||||:  |||||:
727 ProAlaAlaAlaThrAlaThrAlaArgProSerThrLeuProAla.... 741
646 AGATCGGGAATGCGCGCGAAGCTTCAACGGCACTGCAGATATCG.... 691
|||  :::  |||||  |||||  |||||
742 ...ArgAspValLysProLysProLleThrGluAlaLysValAlaGluL 757
692 .....TCAAAACATCATCGCGCGGCGAGGAAATGTCG 727
757 yArgThrSerProSerLysProSerSerAlaProAlaLeuLysProGly 773
728 GCGAGCGGATCGCTGCGAGGATATAAGCGAAGCTCAAAACATTCCTGTC 777
:::  |||  |||||  |||||  :::
774 ProLysThrThrProThrValSerLysAlaThrSerProSerThrLeuVa 790
778 ATGACGGGCTGGGTCTGCTTTCCACCGGAAACAAGATGGCGCGCATCAA 827
```

```
790 lSerThrGly.....ProSerSerArgSerProAlaThr. 801
828 CGATTTGGCAGATATGCGCAACTCAAAAGACTATGCGCGACGAGCCATCC 877
802 .....ThrLeuProLysArgProThr 808
878 GCGATTGGCAGTCCAAACCCCAATGCCGACAAAGGCATAGAAGCCGTC 927
809 SerIleLysThrGluGlyLysProAlaAspValLys..... 820
928 AGCAATATCTTTATGGCAGCATCCCATCAAAAGGATGGAGCTGTCCG 977
820 ..... 820
978 GGGAAATACGCTTGGCGGCATCAGGCACATCTGTCAACGCGTCCG 1027
821 ..ArgMetThrAlaLysSerAlaSerAlaAsp...LeuSerArgSerLys 835
1028 AGATGGCGCGATCGCATTCGGAAGGAAATCCGCGTCACGACAAT 1077
836 ThrThrSerAlaSerSerValLysArgAsnThrThrProThrGlyAlaAl 852
1078 TTGCGCATCGCGCATAG...CAAAATACCCGTCCTCCCTTACCATTCGCG 1124
852 aProProAlaGlyMetThrSerThrArgValLysProMetSerAlaPro 869
1125 AA.....ATATCCGTTCAAACTTGGAGACGCTTACGGCA 1159
869 erArgSerSerGlyAlaLeuSerValAspLysLysProThrSerThrLys 885
1160 AAGAAACATCACCTCCTCAACGCGTCCGCGCTCAACGCGCAAAATGTC 1209
886 ProSerSerAlaProArgValSerArgLeuAlaThr..... 898
1210 AAATGCGACACCAACCCACCGGAGACAGCGGTACCGTTTGACGGTAA 1259
899 .....ThrValS 901
1260 AGGTTTCGAATTTTGAGAG...CACGTGAAATATGATACGAAGCTCG 1306
901 erAla...ProAspLeuLysSerValArgSerLysValGlySerThrGluA 917
1307 ATATTCAGAAATATCGGGGCGGTATACCTAAGGCTAAGCCTGTGTTT 1356
917 snIleLysHisGlnProGlyGly..... 925
1357 GATCGGAACCGGAGATGGGAGTTGTAGGAGCTTAATAAATTGACAAC 1406
926 .....ArgAlaLysValGluLysThrGluAlaAlaThrTh 938
1407 TCGTGACGAGGTGGAGAAAATGTCAGGAAGAGAGAGAGAGTCCAGA 1456
938 rAlaGlyLysProGluProAsnAlaValThrLysAlaAlaGlySerIleA 955
1457 GTAGTCAGTTTAAAGCCCATCGCAACGAGAGATGGGAAAATAAACAAGG 1506
955 lSerAlaGlnLysProProAlaGlyLysValGlnIleValSerLysLys 971
1507 TTAGATTTTAAATCATTTTATAGTGGTGATATCAATAAGAAGGCACAGT 1556
972 ValSerTyrSerHisIleGlnSerLysCysValSerLysAspAsnIle.. 987
1557 AACAGGAGGCATAGTCTAAACCGTGTGTGTACGGGTGATACAAACAA 1606
987 ..... 987
1607 CCTCGCACCTGATAACAT.....GGGTTTATCAAGCGACAGTGGAA 1650
988 .....LysHisValProGlyCysGlyAsnValGlnIleGln 999
1651 ATTAAGAGCCTGATGGAAGTTGGAGTGAACGAAACGAAAGGTGGAA 1700
|||||  |||  |||  ::|||:  |||||  |||||
```









```
38 rGlyInSerThrSerGlyThr.....ThrAsns 49
210 AGCCATCAGTTGGGCCACCTGATGATCAACAGGGCGCGTGAAGNA 259
49 erSerSerProThrThr..... 55
260 ATATCGCTACATATGTCGCTTTCCGATCAGGGCACAAATTCATTCG 309
56 SerProProThrThrSerSerSerProProThrThrSerHisThrSerSe 72
310 CCCTTCG.....ACACCATGCCTCACATTCGGATTCGACGAAGCCGG 353
72 rProSerSerThrThrSerThrGlnSerSerThrAlaAlaThrSerSerS 89
354 TAGTCCGCTGAGCGATTCAGCTTTACCGCA.....TCCATTGGG 394
89 erAlaProSerThrAlaSerSerThrThrSerIleProThrThrSer 105
395 ACGGATACGAACACCATCCCGCGCGGATGATGACGGCCACAGGGCGGC 444
106 ThrGluThrThrThrThrThrProThrAlaSerThrThrProThrTh 122
445 GGCTATCCGCTCCAAAGCGGGAGGATATATACAGCTACAGATATA 494
122 rThrThrAlaAlaProThrThrAlaAlaThrThrThrAlaValThrTh 139
495 AGCGC..... 499
139 laAlaSerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSer 155
500 .....TTGCCCAAAATATCCGCCCTC 519
156 ThrProThrThrThrThrProThrSerThrThrThrThrAlaThrTh 172
520 AACCTGACGACAAACCGGACCGGACGACAAAGCTTGCCGACGTTTCCA 569
172 rThrValProThrThrAlaSerThrThrThrThrThrThrAlaAla 189
570 CAATCGCGCGCTATGCTGACGCAAGGAGTAGCGACGGATTCAACGGC 619
189 hrThrThrAla.....AlaThrThrThrAlaAlaThrThrThrAla 202
620 CCACCGGATACACCCCGAGCTGGACAGATCGGGCAATCCCGCGGAGCC 669
203 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaTh 219
670 TTCAAGGCGACTCAGATATCGTCAAAACATCATCGCGCGCGGAGGAGA 719
219 rThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrTh 236
720 AATTGTCGGCGCAGCGATGCGCGTGCAGGATATAGCGAAGGCTCAACA 769
236 laAlaThrThrThrAlaAlaThrThrThr.....AlaAlaThrThrThr 250
770 TTGCTGTCTGACAGCGCTGGGTCTGCTTCCACCGAAACAGATGGCG 819
251 AlaAlaThrThrThrAlaAlaThrThrThrGlySerProThrSerGly 267
820 CGCATCAGCATTTGGCATATGCGCGCACTCAAGAGACTATGCGCGCAG 868
267 rThrSerThrThrGlyAla.....SerThrSerThrProSerA 280
869 .....CAGCCATCCGCGATT 883
280 laSerThrAlaThrSerAlaThrProThrSerThrSerThrSerAlaAla 296
884 GGGCAGTCCAAAACCCCAATGCGGCACAGGCGATAGAACCGCTCAGCAAT 933
297 AlaThrThrSerThrProThrProThrSerAla...AlaThrSerAlaGl 312
934 ATCTTTATGCGCAG.....C 947
|||
:::
```

```
312 uSerThrThrGluAlaProThrSerThrProThrThrAspThrThrThrp 329
948 CATCCCATCAAAAGGGATGAGCTGTCGGGGAATAACAGGCTTGGCGC 997
|||
:::
329 roSerGluAlaThrThrAlaThrThrSerProGluSerThrThrValSer 345
998 GCATCAGGCGACATCCCTGTCAAGCGGTCCGAGATGGCGCGATCGCAT 1047
|||
:::
346 AlaSerThrThrSerAlaThrThrAlaPheThrThrGluSerHisTh 362
1048 CCGAAAGGGAAT.....CCGCCGTCACGG. 1072
362 rSerProAspSerSerThrGlySerThrSerThrAlaGluProSerSer 379
1073 .....ACAATTTTCCGATCGCGCATACGCCAAATACCCGCTCCCTTACC 1117
379 hrPheThrLeuThrProSerThrAlaThrProSerThrAspGlnPhe... 394
1118 ATCCCGGAATATCCGTTCAACTTGGAGCAGCGTTACGGCAAGAAGAAC 1167
|||
:::
395 .....ThrGlySerSerAlaSerThrGluSerAs 404
1168 ATCACC.....TCCTCAACCGTCGCGCTCAACGCGCAAAATGTCAAAC 1213
|||
:::
404 pSerThrAspSerSerThrValProThrThrGlyThrGluSerIle.... 419
1214 TGGCAGACCAACGCCACCGCAGGCGCTACCGTTTTCAGCGTAAAGGG 1263
:::
420 .....ThrGluSerSerThrThrGluAlaSerThrAsnLeuGly 433
1264 TTTCCGAATTTTGAGAAGCAGCTGAAATATGATACGAGAGCTCGATATCA 1313
:::
434 SerSerThrTyGluSerThrGluAlaLeuGluThrPro...AspGlyAs 449
1314 AGAATTTACGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATCGCA 1363
449 nThrThrSerGlyAsnThrThrProSerProSerPro..... 461
1364 AACCGAGATGGAGGTTTCATAGGAAGCTTAATAAATTCACAACCTCGTGAG 1413
:::
462 .....ArgThrProSerPheAlaAspThrGln 470
1414 CAGTGGAGAAAATGTTTCAGGAACAGCAAGAGAGTCAAGAGTCAAGTCA 1463
|||
:::
471 GlnThrProAspAsnGlyValSerThrGlnHisThrThrIleAsnAspH1 487
1464 GTTTAAAGCCCATCGCAACGAGAAATGGAAATAAAACAGGGTTAGATT 1513
487 sThrThrAlaAsnAlaGlnIlyHis.....A 496
1514 TTAATCATTTTATAGTGGTGATATCAATAGAAAGGCACAGTAAACAGGA 1563
:::
496 laGlyHisHisArgGlyArgAlaGlyGlyArgArgGlySerProGlnGly 512
1564 GGGCATAGTCTAACCCGTGGTGATGTCAGGGTGATACAAACACCTCGGC 1613
|||
513 GlySerHisThrThrProHisProAspArgLeu.....ThrProSe 526
1614 ACCTGAT 1620
526 rProAsp 528
```

seq\_name: SwissProt\_40:AMVB\_PAEPO

seq\_documentation\_block:

ID AMVB\_PAEPO STANDARD; PRT; 1196 AA.

AC P21543;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Beta/alpha-amylase precursor [Includes: Beta-amylase (EC 3.2.1.2);

DE Alpha-amylase (EC 3.2.1.1)]

OS Paenibacillus polymyxa (Bacillus polymyxa).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Paenibacillus.  
 RN [1]  
 RP SEQUENCE OF 1-936 FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-72;  
 RX MEDLINE-87165765; PubMed=2435707;  
 RA Kawazu T., Nakanishi Y., Uozumi N., Sasaki T., Yamagata H.,  
 RA Tsukagoshi N., Uda S.;  
 RA "Cloning and nucleotide sequence of the gene coding for enzymatically  
 RT active fragments of the Bacillus polymyxa beta-amylase.";  
 RT J. Bacteriol. 169:1564-1570(1987).  
 RL [2]  
 RN SEQUENCE OF 689-1196 FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-72;  
 RX MEDLINE-89123046; PubMed=2464578;  
 RA Uozumi N., Sakurai K., Sasaki T., Takekawa S., Yamagata H.,  
 RA Tsukagoshi N., Uda S.;  
 RA "A single gene directs synthesis of a precursor protein with beta-  
 RT and alpha-amylase activities in Bacillus polymyxa.";  
 RL J. Bacteriol. 171:375-382(1989).  
 RN [3]  
 RP SEQUENCE OF 1-776 FROM N.A.  
 RC STRAIN-ATCC 8523;  
 RX MEDLINE-87231094; PubMed=2438660;  
 RA Rhodes C., Strasser J., Friedberg F.;  
 RA "Sequence of an active fragment of B. polymyxa beta amylase.";  
 RL Nucleic Acids Res. 15:3934-3934(1987).  
 RN [4]  
 RP DISULFIDE BOND, AND MUTAGENESIS OF CYSTEINE RESIDUES.  
 RX MEDLINE-91215008; PubMed=1827035;  
 RA Uozumi N., Matsuda T., Tsukagoshi N., Uda S.;  
 RA "Structural and functional roles of cysteine residues of Bacillus  
 RT polymyxa beta-amylase.";  
 RL Biochemistry 30:4594-4599(1991).  
 CC -!- FUNCTION: THE PRECURSOR PROTEIN IS PROTEOLITICALLY CLEAVED TO  
 CC PRODUCE MULTIFORM BETA-AMYLASES AND A 48 Kda ALPHA-AMYLASE AFTER  
 CC SECRETION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in  
 CC polysaccharides so as to remove successive maltose units from the  
 CC non-reducing ends of the chains.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic  
 CC linkages in oligosaccharides and polysaccharides.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO FAMILY 14 OF  
 CC GLYCOSYL HYDROLASES (BETA-AMYLASES).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 13 OF  
 CC GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M15817; AAA85446.1; -;  
 CC EMBL: Y00150; CA668344.1; -;  
 CC PIR: A29130; A29130.  
 CC PIR: A29108; A29108.  
 CC PIR: A32251; A32251.  
 CC HSP: P36924; 1B92.  
 CC InterPro: IPR000461; Alpha\_amylase.  
 CC InterPro: IPR001554; Glyco\_hydro\_14.  
 CC Pfam: PF00128; alpha-amylase; 1.  
 CC Pfam: PF02806; alpha-amylase\_C; 1.  
 CC Pfam: PF01373; Glyco\_hydro\_14; 1.  
 CC PRINTS: PR00750; BETAAMYLASE.  
 CC PROSITE: PS00506; BETA-AMYLASE; 1.  
 CC PROSITE: PS00679; BETA-AMYLASE\_2; 1.  
 CC Multifunctional enzyme; Hydrolase; Glycosidase; Signal;  
 KW Polysaccharide degradation; Repeat.

FT SIGNAL 1 35  
 FT CHAIN 36 1196 BETA/ALPHA-AMYLASE.  
 FT DOMAIN 36 454 BETA-AMYLASE.  
 FT REPEAT 455 558  
 FT REPEAT 565 668  
 FT DOMAIN 669 1196 ALPHA-AMYLASE.  
 FT DISULFID 118 126  
 FT ACT\_SITE 124 124  
 FT ACT\_SITE 198 198  
 FT MUTAGEN 118 118  
 FT MUTAGEN 126 126  
 FT MUTAGEN 358 358  
 FT CONFLICT 1 1  
 FT CONFLICT 67 67  
 FT CONFLICT 100 100  
 FT CONFLICT 154 154  
 FT CONFLICT 177 177  
 FT CONFLICT 227 228  
 FT CONFLICT 330 330  
 FT CONFLICT 425 425  
 FT CONFLICT 493 493  
 FT CONFLICT 532 532  
 FT CONFLICT 559 559  
 FT CONFLICT 665 665  
 FT CONFLICT 681 681  
 FT CONFLICT 686 686  
 FT CONFLICT 725 728  
 FT CONFLICT 736 736  
 FT CONFLICT 741 741  
 FT CONFLICT 758 758  
 SQ SEQUENCE 1196 AA; 130893 MW; A41EA6B70F257064 CRC64;

alignment\_scores:  
 Quality: 133.00 Length: 717  
 Ratio: 0.426 Gaps: 38  
 Percent Similarity: 43.515 Percent Identity: 21.060

alignment\_block:

US-09-303-518D-463 x AMYB\_PAEPO ..

Align seg 1/1 to: AMYB\_PAEPO from: 1 to: 1196

71 CTTACAGATTGGCAACGATCCCTTTATCGGCAGAGTTCTCGACCGTCAG 120  
 |||||:|||||: |||||: |||||:  
 187 ProlysinleTy.....LeuSerGly..... 193  
 121 CATTTGCAACCCGACGGGAAT.....ACCACCTATTGGGACG 158  
 :|||:|||||: |||||: |||||:  
 194 .....GlyProSerGlyGluLeuArgTyProSerTyTyProAlaA 208  
 159 CAGGGGGGAGCTTGCNAGC..... 178  
 ||||| ||||| |||||: |||||:  
 208 laGlyTrpSerTyProGlyArgGlyLysPheGlnAlaTyThrGluThr 224  
 179 .....GCAACGGCCATATCGGAT..... 196  
 |||||:|||||: |||||: |||||:  
 225 AlalysAsnAlaPheArgThrAlaMetAsnAspLysTyTyGlySerLeuAs 241  
 197 .....TGGGAACATACAAAGCCATCA.....GTT 221  
 ||||| ||||| |||||:  
 241 pLysIleAsnAlaAlaTrpGlyThr.LysLeuThrSerLeuSerGlnIle 257  
 222 GGGCCACCTGATGAT.....TCAACAGCGCGCCGTTGAAG 256  
 :|||:|||||: |||||: |||||:  
 258 AsnProProThrAspGlyAspGlyPheTyThrAsnGlyGlyTyAsnSe 274  
 257 GAAATATCGGTACATTGTCGCGCTTTCCGATCACGGGCACAAATTCAT 306  
 ||||| ||||| |||||:  
 274 alaTyTy.Gly.....LysAspPheLeuSerTyTyGln 285  
 307 TCGCCCTTCGACAACCAT.....GCCTCACATTCGCA 338  
 ||||| :|||:|||||: |||||: |||||:





AC p34259;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Atrophin-1 (dentatorubral-pallidoluysian atrophy protein).  
 GN DRPLA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cerebellum, and Brain;  
 RX MEDLINE=95144175; PubMed=7842016;  
 RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,  
 RA Inoue T., Yamada M.;  
 RT "Structure and expression of the gene responsible for the triplet  
 RT repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";  
 RL Nat. Genet. 8:177-182(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96262314; PubMed=8965642;  
 RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,  
 RA Kidwai A.S., Ashworth R.G., Ross C.A.;  
 RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human  
 RT brain.";  
 RL Brain Res. Mol. Brain Res. 36:219-226(1996).  
 RN [3]  
 RP SEQUENCE OF 470-725 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=93315145; PubMed=8325628;  
 RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;  
 RT "Novel triplet repeat containing genes in human brain: cloning,  
 RT expression, and length polymorphisms.";  
 RL Genomics 16:572-579(1993).  
 CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,  
 CC OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE  
 CC LIVER, THYMUS AND LEUKOCYTES.  
 CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC  
 CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO  
 CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN  
 CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE  
 CC DISEASE.  
 CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-  
 CC PALLIDOLUSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE  
 CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE  
 CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUY'S BODY. CLINICAL FEATURES  
 CC ARE MYOCLONUS, EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF  
 CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH  
 CC IN THE FOURTH.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; D31840; BAA06626.1; -;  
 CC ENBL; U23851; AAB50276.1; -;  
 CC EMBL; L10377; -; NOT\_ANNOTATED\_CDS.  
 CC HSSP; P00651; 1LRA.  
 CC MIM; 125370;  
 CC InterPro; IPR002951; Atrophin.  
 CC PRINTS; PR01222; ATROPHIN.  
 KW Triplet repeat expansion; Polymorphism.  
 FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).  
 FT DOMAIN 302 305 POLY-PRO.  
 FT DOMAIN 376 382 POLY-SER.  
 FT DOMAIN 386 397 POLY-SER.  
 FT DOMAIN 442 447 POLY-PRO.  
 FT DOMAIN 479 483 POLY-HIS.

FT	DOMAIN	484	497	POLY-GLN.
FT	DOMAIN	504	507	POLY-PRO.
FT	DOMAIN	564	574	POLY-SER.
FT	DOMAIN	704	707	POLY-PRO.
FT	DOMAIN	802	815	ARG/ALA-RICH (MIXED CHARGE).
FT	DOMAIN	816	827	ARG/GLU-RICH (MIXED CHARGE).
FT	DOMAIN	925	934	ARG/GLU-RICH (MIXED CHARGE).
FT	CONFLICT	94	94	MISSING (IN REF. 2).
FT	CONFLICT	333	333	Y -> H (IN REF. 2).
FT	CONFLICT	339	339	M -> I (IN REF. 2).
FT	CONFLICT	541	541	P -> T (IN REF. 3).
FT	CONFLICT	1028	1028	G -> A (IN REF. 2).
SQ	SEQUENCE	1185 AA;	124785 MW;	56C306267331C005 CRC64;

alignment\_scores:  
 Quality: 132.00 Length: 458  
 Ratio: 0.714 Gaps: 22  
 Percent Similarity: 40.393 Percent Identity: 22.052

alignment\_block:  
 US-09-303-518d-463 x DRPL\_HUMAN ..

Align seg 1/1 to: DRPL\_HUMAN from: 1 to: 1185

65	CACACGCCCTCAGATTGGCAACAGATCCCTTTATCCGCGCAGGTTCCTCGAC	114
253	HisProPro.....ProThrThrProIleSerValSerSerSerG1	266
115	CGTCAGCATTTTGAACCCGCGGAAATACACCTATTTCGCGCAGCAGGGG	164
266	YAlaSerGlyAlaProProThrThrProThrThrProValGlyGlyG	283
165	GGAGCTTGCNAGCGCAACGCCATATCGATTGGGAAACATACAAAGCC	214
283	LYAsnLeuProSerAlaProProAlaAsnPhe.ProHisValThrPr	299
215	ATCAGTTGGCGCACCTGATTTCAACAGCGCGCTTGAAGGAATAATC	264
299	OsanLeu...ProProProAlaLeu...ArgProLeuAsnAlase	314
265	GGCTACA.....TTCTCCGCTTTCCGATCACGGGC.....	295
314	rAlaSerProGlyLeuGlyAlaGlnProLeuProGlyHisLeuProS	331
295	.....	295
331	erProTyrAlaMetGlyGlnGlyMetGlyGlyLeuProProGlyProGlu	347
296	.....ACAAATTCATTCGCGCTTCGACACAC	321
348	LysGlyProThrLeuAlaProSerProHisSerLeuProProAlaSerSe	364
322	CATGCGCTCATTCCTCCAGTTCGACGAAGCCGGTAGTCGCGTTGACGGATT	371
364	rSerAlaProAlaProProMetArgPheProTyrSerSerSerSerS	381
372	CAGCCTTTACCGCATCCATTGGGCGGATACGACACACCATCCCGCGGAGC	421
381	erSerAlaAlaAlaSerSerSerSerSerSer.....	393
422	GCTATGACGGCCACAGGGCGCGGTATCCGCTCCCAAGCGCGGAGG	471
394	.....SerSerAlaSerProPheProAlaSer.....	402
472	GATATATACAGTACGACATAAAGCGTTGCCAAATATATCCGCTCAA	521
403	.....GlnAlaLeuProSerTyrProHisSerP	412
522	CTGACCGCACACCGGACCGGACACAGCGGCTTCCGACCGCTTCCACA	571
412	heProProProThrSerLeuSerValSerAsnGlnProProPolysTyrThr	428

```
572 ATGCGCGCGCTATGCTGACCAAGGAGTAGGCGACGGATTCAAACGCGCC 621
    :::::
429 GlnProSerLeu..... 432
622 ACCGATACAGCCGAGCTGGACAGATGGGCAATGCGCGGAGCGCTT 671
    :::::
433 .ProSerGlnAla...ValTrpSerGlnGlyProProProProt 448
672 CAACGCGACTGCAGATATGCTCAAAACATCATCGCGCGCGCAGAGAA 721
    :::::
448 y-GlyArgLeuLeuAlaSer..... 455
722 TTGTGCGCGCAGCGATGCGTCGAGGTATTAAGCAAGGCTCAACATTT 771
    :::::
456 ...AsnAlaHisProGlyProPheProProSerThrGlyAlaGlnSer.T 471
772 GGTGTCATCAGCGCTGGTGTCTTCCACGCAAAACAGATGGCGCG 821
    :::::
471 hrAlaHisProProValSerThrHisHisHisHisGlnGlnGln 487
822 CATCAACGATTGGCAGATATGCGCAACTCAAGACTATGCGCGCAGCAG 871
    :::::
488 GlnGlnGln.....GlnGlnGlnGlnGlnGlnGln 498
872 CCATCGCGGATGGGAGTCCAAACCCCAATGCGCGCAGCAGCATAGAA 921
    :::::
498 sHisGlyAsnSerGlyProProPro..... 507
922 GCGTCAGCAATATCTTTATGCGAGCATCCCATCAAGGATGGAGC 971
    :::::
508 .....GlyAlaPheProHisProLeuGlnGlyGly 517
972 TGTGCGGGAAATACGGCTGGCGGCATCACGCACATCT..... 1014
518 .....SerSerHisHisAlaHisProTyrAlaM 527
1014 ..... 1014
527 etSerProSerLeuGlySerLeuArgProTyrProProGlyProAlaHis 543
1015 .....GTCACGCGTCCGAGATGGCGCGCATGCC 1043
544 LeuProProHisSerGlnValSerTyrSerGlnAlaGlyProAsnG1 560
1044 ATTGCGGAAAGAAATCGGCC...GTCACGCAATATTTGCCGATGCGG 1090
560 yProProValSerSerSerSerSerSerSerSerSerSerSerGlnGlyS 577
1091 CATAC.....GCCAAA 1101
577 etTyrProCysSerHisProSerProSerGlnGlyProGlnGlyAlaPro 593
1102 TACCGTCCCTTACCATTCGCAATA.....TCCGTCAACCTGGA 1145
594 TyrProPhePro.ProValProThrValThrSerSerAlaThrLeuS 610
1146 GCACGCTTACGGCA.....AAGAAACAA 1168
610 etThrValIleAlaThrValAlaSerSerProAlaGlyTyrLysThrAla 626
1169 TCACCTCTCAACCG 1183
627 SerProGlyPro 631
```

seq\_name: SwissProt\_40:GTFB\_STRMU

seq\_documentation\_block:

ID GTFB\_STRMU STANDARD; PRT; 1476 AA.

AC P08987; O69381; O69384; O69387; O69390; O69396;  
DT 01-NOV-1988 (Rel. 09, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)

```
DE (Sucrose 6-glucosyltransferase).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RN PEMS Microbiol. Lett. 161:331-336(1998).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS. THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl](N) = D-
CC fructose + [(1,6)-alpha-D-glucosyl](N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M17361; AAA8588.1; -
CC EMBL: D88651; BAA26101.1; -
CC EMBL: D88654; BAA26105.1; -
CC EMBL: D88657; BAA26109.1; -
CC EMBL: D88660; BAA26113.1; -
CC EMBL: D89977; BAA26119.1; -
CC PIR: B33135; B33135.
CC InterPro: IPR002479; CW_binding.
CC InterPro: IPR003318; Glyco_hydro_70.
CC Pfam: PF01473; CW_binding_1; 13.
CC Pfam: PF02324; Glyco_hydro_70; 1.
CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 1476 GLUCOSYLTRANSFERASE-I.
FT DOMAIN 35 1051 CATALYTIC (APPROXIMATE).
FT DOMAIN 1097 1476 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1097 1130 A REPEAT.
FT DOMAIN 1161 1470 5 X TANDEM REPEATS.
FT REPEAT 1161 1210 1.
FT REPEAT 1225 1275 2.
FT REPEAT 1290 1340 3.
FT REPEAT 1355 1405 4.
FT REPEAT 1420 1470 5.
FT REPEAT 62 62
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
```

```
FT VARIAT 86 86 S -> I (IN STRAIN MT4239).
FT VARIAT 89 89 S -> F (IN STRAIN MT4251).
FT VARIAT 168 168 K -> N (IN STRAIN MT4251).
FT VARIAT 276 276 D -> S (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIAT 399 399 N -> R (IN STRAIN MT4239).
FT VARIAT 474 474 I -> T (IN STRAIN MT4239).
FT VARIAT 512 512 K -> R (IN STRAIN MT8148).
FT VARIAT 519 519 F -> Y (IN STRAIN MT8148).
FT VARIAT 701 701 T -> I (IN STRAIN MT8148).
FT VARIAT 708 708 A -> V (IN STRAIN MT8148).
FT VARIAT 938 938 F -> L (IN STRAIN MT8148).
FT VARIAT 952 957 YGTPVA -> FGKPEV (IN STRAINS MT4245,
MT4251 AND MT8148).
FT VARIAT 963 964 NT -> SV (IN STRAINS MT4245, MT4251 AND
MT8148).
FT VARIAT 968 970 VDG -> ADS (IN STRAINS MT4245, MT4251 AND
MT8148).
FT VARIAT 1086 1086 A -> T (IN STRAIN MT4239).
FT VARIAT 1158 1158 S -> N (IN STRAIN MT4239).
FT VARIAT 1163 1163 H -> Y (IN STRAIN MT4251).
FT VARIAT 1168 1168 E -> K (IN STRAIN MT8148).
FT VARIAT 1182 1182 Y -> C (IN STRAIN MT8148).
FT VARIAT 1234 1234 A -> P (IN STRAIN MT4239).
FT VARIAT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIAT 1264 1264 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIAT 1272 1272 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIAT 1329 1329 Y -> H (IN STRAINS MT4239, MT4245, MT4251
AND MT8148).
FT VARIAT 1394 1394 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIAT 1402 1402 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIAT 1459 1459 Y -> H (IN STRAIN MT4467).
FT CONFLICT 570 570 R -> A (IN REF. 1).
FT CONFLICT 800 817 ADDVRVAASTAPSTDGK -> LIKFMALRLARPHQOMA
(IN REF. 1).
FT CONFLICT 1310 1310 H -> L (IN REF. 1).
SQ SEQUENCE 1476 AA; 165685 MW; 3479BG2B07694D98 CRC64;
```

## alignment\_scores:

Quality: 129.50 Length: 751  
Ratio: 0.415 Gaps: 46  
Percent Similarity: 41.545 Percent Identity: 21.571

## alignment\_block:

US-09-303-518d-463 x GTFB\_STRMU

Align seg 1/1 to: GTFB\_STRMU from: 1 to: 1476

```
64 GCACAGCCTCAGATTGGCAACGATCCCTTTATCCGGCAGGTCTCGA 113
.....
617 SerTyrAlaLeuLeuLeuThrAsnLysSerValProArgValTyrTy 633
.....
114 CGTCAGCATTTCGACCCGACGCGGAAATAC..... 144
...
633 rGlyAspMetPheThrAspGlyGlnTyrMetAlaHisLysThrIleA 650
.....
145 .....CACCTATTGGCAGCAGG..... 162
.....
650 snTyrGluAlaIleGluThrLeuLeuLysAlaArgIleLysTyrValSer 666
.....
163 GGGAGGCTGCNAGCGCAACGGCCATATC..... 192
.....
667 GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGluIleIleTh 683
.....
193 .....GGATTGGGAACATACAAAGCCATCAGTTGGCCACC 229
...
683 rSerValArgTyrGlyLysGlyAlaLeuLysAlaThrAspThrGlyAspA 700
.....
230 TGATGATTCAACAG.....GGGCCGTTGAAGAAATATCGGTAC 270
...
.....
```

```
700 rgThrThrArgThrSerGlyValAlaValIleGluGlyAsnAsnProSer 716
271 ATT...GTCGCGCTTTCCGAT.....CA 290
...
717 LeuArgLeuLysAlaSerAspArgValValAsnMetGlyAlaAlaH1 733
291 CGGGCAACAATTCATTCCGCTTC.....GACAAC.....C 322
...
733 sLysAsnGlnAlaTyrArgProLeuLeuLeuThrAspAsnGlyIleL 750
323 ATGCTCACAATTCGATTCTGACGAACCGGTAGTCCCTGTGACGGATT 372
...
750 ysAlaTyrHisSerAspGlnGluAlaGly..... 760
373 AGCCTTACCGCATTCATTGGGACGGATACGAA.....CACCATCCCGC 416
...
761 ...LeuValArgTyrThrAsnAspArgGlyGluLeuIlePheThrAlaA 776
417 CGAC.....GGCTATGACGGGCACAGGGCGGCTATCCC..... 453
...
776 asPileLysGlyTyrAlaAsnProGlnValSerGlyTyrLeuGlyVal 793
454 ..GCTCCCAAGCGCGAGGATATATACAGCTACGACATAAAGCGGTT 501
...
793 rpValProValGlyAla.....AlaAla 800
502 GCCCAAAATATCCGCTCAACCTGACCGACACCGGACGACCGGACAACG 551
...
801 AspGlnAspValArgValAlaAlaSerThrAlaProSerThrAspGly 817
552 GCTTGGCGCGCTTTCACAAATCGCGCGCTATGCTGACGCAAGGAGTAG 601
...
817 s.....SerValHisGlnAsnAlaAlaLeuAspSerArgValMetP 831
602 GCGACGGATTCAACCGCGCCCGATACAGCCCGAGCTGGACAGATCG 651
...
831 heGluGlyPheSerAsnGlnAlaPheAlaThrLysLysGluGlyTyr 847
652 GCGAATGCCCGCGGAGCTTCAAC.....GGCAC 680
...
848 ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluTrpGly 864
681 TGCAGATATCGTCAAAAACATCATCGCGCGCGAGGAGAAATGTCGCG 730
...
864 lThrAsp.....PheGluMetAlaProGlnTyrValSerS 876
731 CAGCGGATGCCGTG.....CAGGGTATAAGCAAGCTCAAAACATT 771
...
876 erThrAspGlySerPheLeuAspSerValIleGlnAsnGlyTyrAlaP 892
772 GCTGTATGCACCGCTTGGTCTGCTTTCACCGAAAAACAAGATGGCGG 821
...
893 ThrAspArgTyrAspLeuGly...IleSerLysProAsnLysTyrGly 908
822 CATCAAGATTTCGAGAT...ATGGCGCAACTCAAAGACTATGCGCGAG 868
...
908 rAlaAspAspLeuValLysAlaIleLysAlaLeuHisSerLysGlyIle 925
869 CAGCCATCCCGGATGGGACGTCCAAACCCCAATGCCGCAAGGCATA 918
...
925 ysValMetAlaAspTrpValProAspGln..... 934
919 GAAGCGCTACGCAATATCTTTATGGCAGCATCCCATCAAGGG...AT 965
...
935 .....MetTyrAlaPheProGluLysGluValVa 944
966 TGGAGCTGTCCGG...GGAAATACGCTTGGGCGGCATCAGCGCACATC 1012
...
944 lThrAlaThrArgValAspLysTyrGlyThr.....P 955
1013 CTGCTAAGCGGTGCGAGATG...GGCGGATCGATTGCCGAAGGAA 1059
...
955 roValAlaGlySerGlnIleLysAsnThrThrValValAspGlyLys 971
```









RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;  
 RX MEDLINE=96081227; PubMed=8541849;

RA Schmitt I., Epplen J.T., Riess O.,  
 RT "Dominant neuronal expression of the gene responsible for  
 RL dentatubular-pallidolusian atrophy (DRPLA) in rat.";  
 Hum. Mol. Genet. 4:1619-1624 (1995).

CC -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH

CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.  
 CC -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES  
 CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; U31777; AAA80337.1; -;

DR EMBL; X89453; CAA61623.1; -;

DR InterPro; IPR002951; Atrophin.

DR PRINTS; PR01222; ATROPHIN.

FT DOMAIN 165 171 POLY-PRO.

FT DOMAIN 303 306 POLY-PRO.

FT DOMAIN 377 383 POLY-SER.

FT DOMAIN 387 391 POLY-SER.

FT DOMAIN 440 446 POLY-PRO.

FT DOMAIN 477 480 POLY-HIS.

FT DOMAIN 481 489 POLY-GLN.

FT DOMAIN 502 505 POLY-PRO.

FT DOMAIN 562 572 POLY-SER.

FT DOMAIN 702 705 POLY-PRO.

FT CONFLICT 455 455 N -> S (IN REF. 2).

FT CONFLICT 594 594 F -> L (IN REF. 2).

FT CONFLICT 689 689 P -> R (IN REF. 2).

FT CONFLICT 717 717 T -> M (IN REF. 2).

FT CONFLICT 737 737 A -> V (IN REF. 2).

FT CONFLICT 965 965 MISSING (IN REF. 2).

SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9BIF CRC64;

alignment\_scores:

Quality: 126.50 Length: 509

Ratio: 0.591

Percent Similarity: 42.043 Percent Identity: 24.165

alignment\_block:

US-09-303-518D-463 x DRPLA\_RAT ..

Align seg 1/1 to: DRPLA\_RAT from: 1 to: 1183

12 CCGCAAAATATCCCTTATTCGTCCATACATGCG.....AG 46

171 ProAspSerIleProArgGlnProGluSerGlyPheGluProHisProSe 187

47 TGTGCTGCTCCGATGCA.....TGACACGCC 72

187 rValProProThrGlyTyrlHisAlaProMetGluProProThrSerArgL 204

73 TCAGATTGGCAACAGCATCCCTTTATCCGCGCAGGTCTTCGACCG..... 116

204 euPheGlnGlyPro.....ProProGlyAlaProProHis 216

117 .....TCAGATTTCGAAACCGGCGGGAATACCA 148

217 ProGlnLeuTyProGlySerAlaGlyGlyValLeuSerGlyProPr 233

149 TATTTCGGCAGCAGGGGAGCTTCCNAGCGCAA...CGCCCATATCGGA 195

233 oMetGlyProGlyGlyAlaAlaAlaSerSerValGlyProProSerG 250

196 TTGGGAAA...CATACAAAGCCATCAGTTGGGCCACCTGATGATCA 242

250 lyGlyLysGlnHisProProThrThrProIleProIleSerSerSer 266  
 243 GCGCG.....CGTTGAAGAAATATCGCTACATGTCCGC 279  
 267 GlyAlaSerGlyAlaProProAlaLysProProAsnThr..... 279  
 280 TTTTCCGATCACGGGCACAAAT.....TCCATTCCCTTCGACAAACA 323  
 280 ....ProValGlyAlaGlyAsnLeuProSerAlaProProAlaThrP 295  
 324 TGCTCACAATT...CGATTCTCACGAAGCG.....GTAGTCCCGTTG 364  
 295 heProHisValThrProAsnLeuProProProAlaLeuArgProLeu 311  
 365 ACGGATTACGCTTTACCGCATCCATTGGGACGGATACGAACACCATCC 414  
 312 ...AsnAsnAlaSerAlaSerProGlyMetGlyAlaGlnProIlePr 327  
 415 G.....CCGACGGCTATGACGGCCACAGCGCGCGGCTA 449  
 327 oGlyHisLeuProSerProHisAlaMet...GlyGlnGlyMetSerGlyL 343  
 450 TCCCGCTCCCAAAGCGCGAGGATATATACAGCTACGACATAAAGCG 499  
 343 euProProGlyProGlyLysGlyProThrLeuAlaProSerProHisPr 359  
 500 TTGCCCAAAATATCCGCTCAAC.....TGACCGAC 531  
 360 LeuProAlaSerSerSerAlaProGlyProProMetArgTyProTy 376  
 532 AACCGCAGCAGCGACACGGCTTCCGACCGTTTCCACAATCCCGCGC 581  
 376 rSerSerCysSerSerSerValAlaAlaSerSerSerSerAla. 392  
 582 TATGCTGACGCAAGGATAGGACGATTCAACCGCGCCACCCGATACA 631  
 393 .....AlaThrSerGlnTyProAlaSerGlnThr 402  
 632 GCGCCGAGCTGGACACATCGGCGATGCGCGGACCGCTTCAACGCGCA 681  
 403 LeuProSerTyProHisSerPheProProPro.....ThrSerMe 416  
 682 GCAGATATCGTCAAAACATCATCGCGCGGAGAGAAATGTGCGCGC 731  
 416 tSerValSerAsnGlnProProLysTyThrGlnProSerLeuProSerG 433  
 732 AGCGGATGCGGTGCAGGGTATAAGCGAAGGCTCAACATTGCTGTCATGC 781  
 433 InAlaVal..... 435  
 782 ACGGCTTGGTCTGCTTCCACCGAAACAAAGATGCGCGCATCAACGAT 831  
 436 .....TrpSerGlnGlyProProPro.....ProProProTy 447  
 832 TTGGCAGATATGCGCAACTCAAAAGACTATCGCGCAGCAG...CCATCG 878  
 447 rGlyArgLeuLeuProAsnAsnThrHisProGlyProPheProPro 464  
 879 CGATTGGGCGAGTCCAAACCC.....CCAATCCGCGACAAAGCATAGA 922  
 464 hrGlyGlyGlnSerThrAlaHisProProAlaProAlaHisHisHis 480  
 923 CGTTCAGCAATATCTT..... 938  
 480 sGlnGlnGlnGlnGlnProGlnProGlnProGlnProGlnHisHisH 497  
 939 .....TATGCGACGCATCCCATCAAAAGGAT 965  
 497 isGlyAsnSerGlyProProProProGlyAlaTyProHisProLeuGlu 513  
 966 TGGAGCTGTCCGGGAAATACGGCTTGGCGGCATCAGCGCACATCCT. 1014



```

340 TCTGACGAAGCGGTAGTCCGCTGACGATTCAGCCTTTACCGCATCA 389
: |||||: : |||||: |||||: |||||:
169 l...ThrSerThrLeuSerSerThrThrSerSerAsnProThrThrS 185
: |||||: : |||||: |||||: |||||:
390 TTGGGAGGATAGACACATCCCGCGAGGCTATGACGGGCGACAGG 439
: |||||: |||||: |||||: |||||:
185 erLeuSerSerThrThrSerSerSerThrThrSerThrProSer 201
: |||||: |||||: |||||: |||||:
440 GCGGCGCTATCCCGCTCCCAAGGCGGAGGATATACAGCTACGAC 489
: |||||: |||||: |||||: |||||:
202 SerThrSer.....ThrSerSerSerThrThrSerSerSe 215
: |||||: |||||: |||||: |||||:
490 ATAAAGCGCTTCCCAAAATATCCCGCTCAACCTGACCGACAGCGAG 539
: |||||: |||||: |||||: |||||:
215 rSerThrSerThrSerSerSerThrThrSerThrProSerThrS 232
: |||||: |||||: |||||: |||||:
540 CACGGGACAGCGCTTCCCGACCGTTCCCAATGCGGCGCTATCTGA 589
: |||||: |||||: |||||: |||||:
232 erThrSerSerSerLeuThrThrSerSerSerSerThrThrSer 248
: |||||: |||||: |||||: |||||:
590 CGCAAGGAGTAGCGGCGGATCAAGCGCGCCACCGCATACAGCCCGAG 639
: |||||: |||||: |||||: |||||:
249 GlnSerSerThrThrSerSerSerSer.....ThrSerThrSerProSe 264
: |||||: |||||: |||||: |||||:
640 CTGACAGATCGGCGATCGCGGCGGAGCCCTTCAAGCGGCACTGCATAT 689
: |||||: |||||: |||||: |||||:
264 rSerThrSerThrSer.....SerSerSerThrThrSerThrProS 278
: |||||: |||||: |||||: |||||:
690 CGTCAAAACATCATCGGCGGCGGAGGAAATGTCGGCGGAGGCGATG 739
: |||||: |||||: |||||: |||||:
278 erSerLeuSerThrSerAlaSerSerThrThrSerSerThrSerThr 294
: |||||: |||||: |||||: |||||:
740 CGGTGACGGGTATAGCGAGGCTCAACATGCTGTCATGCGCGGCTTG 789
: |||||: |||||: |||||: |||||:
295 .....SerThrSerProSerLeuThrThrSerSerSerProTh 306
: |||||: |||||: |||||: |||||:
790 GTCTGCTTTCACCGGAAACAGATGCGCGCATCAACGATTTGGCAGA 839
: |||||: |||||: |||||: |||||:
306 rLeu.....AlaSerThrSerProSerS 314
: |||||: |||||: |||||: |||||:
840 TATGGCGCAACTCAAGACTATCGCGAGCAGCAGCCATCGCGATTTGGCAG 889
: |||||: |||||: |||||: |||||:
314 erThrSerLeuSerSerThrThrSerSerSerSerSerLeuGlySer 330
: |||||: |||||: |||||: |||||:
890 TCAAAACCCCAATCGCGCACAGGATAGAGCGGTCACCAATATCTTT 939
: |||||: |||||: |||||: |||||:
331 Ser.....IleAlaSerSerSerThrSerValSerLe 341
: |||||: |||||: |||||: |||||:
940 ATGGGAGCATCC...CCATCAAGGGATGAGCTGTCGGGGGAAATA 986
: |||||: |||||: |||||: |||||:
341 uThrSerProSerThrProValThrSerValProSerThrSerSerAsnV 358
: |||||: |||||: |||||: |||||:
987 GCGCTTTGGCGGCATCACGGCACATCTGTCTCAAGCGGTCGACATGGGCG 1036
: |||||: |||||: |||||: |||||:
358 aAla..... 359
: |||||: |||||: |||||: |||||:
1037 CGATCGCATTTGCGAAGGAATCCGCGCTCAGCGACAATTTTCCCGAT 1086
: |||||: |||||: |||||: |||||:
360 .....ThrProSerMetThrSerSerThrVa 368
: |||||: |||||: |||||: |||||:
1087 GCGGCATACGCCAAATACCCGT..... 1108
: |||||: |||||: |||||: |||||:
368 iGluThrThrValSerSerGlnSerSerSerSerSerSerSerSerS 385
: |||||: |||||: |||||: |||||:
1109 .....CCCTTACCATTCGCCAAATATCCGTTCAAACTTG..... 1144
: |||||: |||||: |||||: |||||:
385 erileSerThrThrileProSerPheSerMetSerThrThrPheThrThr 401
: |||||: |||||: |||||: |||||:
1145 ...AGACGCTTACGCAAGAAACATCA...CCTCCTCAACCGTGGCG 1188
: |||||: |||||: |||||: |||||:
402 ValSerGlyValThrThrMetThrThrThrTrpCysProThrSerSerGI 418
: |||||: |||||: |||||: |||||:

```

```

1189 CCGTCAACGCGCAAAATGCTCAAACTGGCAGACCAACGCCACC 1231
: |||||: |||||: |||||: |||||:
418 userGluThrSerThrLeuThrSerMetHisGluThrValThr 432
: |||||: |||||: |||||: |||||:
seq_name: SwissProt_40:HIG_DROME
seq_documentation_block:
ID HIG_DROME STANDARD; PRT; 958 AA.
AC Q09101;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Locomotion-related protein Hikaru genki precursor.
GN HIG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC ISSUE-Head;3498; PubMed-8461133;
RX Hoshino M., Matsuzaki F., Nabeshima Y.-I., Hama C.;
RA "Hikaru genki, a CNS-specific gene identified by abnormal locomotion
RT in Drosophila, encodes a novel type of protein.";
RL Neuron 10:395-407(1993).
CC -1- FUNCTION: HAS A ROLE IN THE DEVELOPMENT OF CNS FUNCTIONS INVOLVED
IN LOCOMOTOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN PCC NEURONS AND NEUROBLASTS
IN THE PROCEPHALIC NEUROGENIC REGION IN THE CENTRAL NERVOUS
SYSTEM.
CC -1- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING AND/OR AFTER NEURONAL
DIFFERENTIATION AND DURING CELL SPECIFICATION OR AXOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D13884; BAA02984.1; -
EMBL; D13885; BAA02985.1; -
EMBL; D13886; BAA02986.1; -
EMBL; D13887; BAA02987.1; -
HSSP; P10998; 1VVD
Flybase; FBgn0010114; hig.
InterPro; IPR003599; Ig.
InterPro; IPR003006; Ig_MHC.
InterPro; IPR000436; Sushi_SCR_CCP.
Pfam; PF00047; Ig; 1.
Pfam; PF00084; sushi; 5.
SMART; SM00032; CCP; 5.
SMART; SM00409; IG; 1.
Glycoprotein; Alternative splicing; Immunoglobulin domain; Repeat;
Sushi; Signal.
FT SIGNAL 1 31
FT CHAIN 32 958
FT DOMAIN 630 709
FT DOMAIN 713 769
FT DOMAIN 772 828
FT DOMAIN 831 891
FT DOMAIN 893 952
FT SITE 318 320
FT SITE 318 320
FT DISULFID 714 755
POTENTIAL.
LOCOMOTION-RELATED PROTEIN HIKARU GENKI.
IG-LIKE C2-TYPE DOMAIN.
SUSHI 1.
SUSHI 2.
SUSHI 3.
SUSHI 4.
CELL ATTACHMENT SITE.
BY SIMILARITY.

```

487 GACATAAAGGCGTTGCCAAAATATCCGCTCAACCTGACCGACAACGC 536  
 139 ASPhisHisGlyAspProGluAspLysArgLeuAsnAla.....AsnAr 153  
 537 CAGCACCGGACAAACGG.....CTTGGCGACCGTTTCC 568  
 153 glysGlnGlyLysArgArgAlaGlySerGlyArgArgArgIleG 170  
 569 ACAATCCGCGGCGCTATGCTGACGCAAGCAGTAGTGGCAGCATGTTCAAACGC 618  
 170 luAsnGluAsnGlyGlnThrGlyArgGlyArgGlySerArgTyrLysArg 186  
 619 .....GCCACCGATACAGCCCGACTGGACAGATCCGG 653  
 187 HisAlaIleLeuHisAspThrGluLaserProGluThrAspArgTirAl 203  
 654 CAATGCCGCGAAGCCTTCAACGC.....ACTGCAG 685  
 203 aglySerLysLeuAlaGluGlyAspValTyrValHisIleAla 220  
 686 ATATCGTCAAAACATCATCGGCGCGCAGGAG..... 718  
 220 spIleLeu.LysSerArgGluProAsnArgGluLeuLysSerLysLeuHi 236  
 719 .....AAATGTGCGCGCAGCGATGCGTGCAGGGTATAAGCGAAGG 761  
 236 slsLysLeuLysMetLysAlaArgLeuAsnLysCysLeuAlaGluGlyL 253  
 762 CTCAAACATTCTCATGCACCGGTTGGGTGCTTTCACCGGAAACA 811  
 253 ysglu.....LysCysThrArgLeuLeuLysLysLysProLys.... 265  
 812 AGATGGCGCGCATCAACGATTTGGCAGATATGCGCAACTCAAGACTAT 861  
 266 .....LysLysValVa 269  
 862 GCCGACAGCCATCCGCGATTGGCGAGTCCAAACCCCATGCGCGACA 911  
 269 igLulysGluGlnThrLeuLysLysLysLysLysPheProLysGluGlu 286  
 912 AGGCATAGAAGCCGTGACGAATATCTTTATGCACCCATCCCATCAAG 961  
 286 lnSer...LysGluLysValLeuLysGlnGlyInThrProLysGluasp 301  
 962 GGATTGGAGCTGTCGCGGAAAATACGGCTTGGCGCGCATCAC..... 1004

```
591 laProhMetProSerCysValProThrThrValLeuThrAsnTyrSer 607
1603 CAACCTCGGCACCTGATAAATGCGGTTTATCAAGCAGCAGTGAAT 1652
608 GluAspSerAlaPro.....SerIleArgil 616
1653 TAAAGACCTGATGGAGTGGAGCTGAAACGAAAGAGTGGGAAG 1702
616 eluYIlePheAsnGlySerHisSerPheGlu...ProSerGlyValMetA 632
1703 TGATGACCAACGACACATGTTCCCAAAAGATTGGGATGAGCTAGAAT 1752
632 laValProProHisSerThrValLeuMetAspCysMetTyrProArgVal 648
1753 AGGCTGAAGTACTTCGCTTGGGAAAGTAGAATAATGCTTAAAGATAA 1802
649 ArgGlyThrProGluTrpSerTrpThrSerTrpTyrMetGlnTyrSerTh 665
1803 TAAATGG 1809
665 rGlyTrp 667
seq_name: SwissProt_40:YQ03_CAEEL
seq_documentation_block:
ID YQ03_CAEEL STANDARD; PRT; 1251 AA.
AC Q09350;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 133.5 kDa protein F26C11.3 in chromosome II.
GN F26C11.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Matthews P.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z47072; CAAB7369.1; -
DR WormPep; F26C11.3; CE01561.
DR InterPro; IPR000436; Sushi_SCR_OCP.
DR Pfam; PF00084; sushi; 1.
DR SMART; SM00032; CCP; 1.
KW Hypothetical protein.
FT DOMAIN 120 414 SER/THR-RICH
SQ SEQUENCE 1251 AA; 133498 MW; 2B959ECA03B9954A CRC64;

alignment_scores:
Quality: 124.50 Length: 413
Ratio: 0.635 Gaps: 18
Percent Similarity: 47.458 Percent Identity: 20.339

alignment_block:
US-09-303-518d-463 x YQ03_CAEEL
..
Align seg 1/1 to: YQ03_CAEEL from: 1 to: 1251

140 AATACCACCATATTCGCGACGAGGGGAGCTTGCNACGCAAGCGGCAT 189
||| :::::::::::::::::::: |||
208 AsnLeuSerThrThrThrSerSerSerThrMetLeuSerSerThrThrLe 224
```

```
190 ATCGGATTGGGAAACATACAAAGCCATCAGTTGGCCACCTGATGATTC 239
224 uLeuThrThrGluThr..... 229
240 ACAGCGCGCCGTTGAAGGAAATATCGGCTACATTTGCGGTTTCCGATC 289
230 .....GluThrArgGluSerSerSerThrGlySerThr.....Gln 241
290 ACGGCAAAATTCATTCGCGCTTCGACAAACCATGCTCACAATTCGGAT 339
242 ThrThrThrProSerThrGluProSerThrThrIle..... 253
340 TCTGACGAAGCGGTAGTCCGCTTACGAGATTACGCTTTTACCAGATCCA 389
254 .....ThrThrProMetGluGlnSerSerThrValSerSerV 266
390 TT.....GGGACGGATACGAACACCATC 412
266 alGlnLysThrArgThrSerGluAspLysProSerSerSerThrThrVal 282
413 CGCGCGCGGCTATACGCGGCACAGGCGGCGCTATCCGCTCCCAAA 462
283 ProThrSerAlaSerThrSerGluSerSerSerProMetAlaG1 299
463 GCGCGAGGATATATACAGCTACGACATAAAGCGTTGCCCAAAATAT 512
299 uThrSerSer.....SerSerThrThrSerGlnSerSerPro...AlaS 313
513 CGCGCTCAACCTGACCGACAC.....GCAGCACCGGACAC 550
313 erThrSerThrValProGluSerSerThrValGlySerThrProThrThr 329
551 GCTTGGCGGCGCTTCCACAATGCCGCGCTATGCTGACGCAAGGAGTA 600
330 GlyLeuThrThrLeuSerThr.....AsnGluG1 339
601 GCGCAGGATTCAACGCGCCACCGATACAGCCCGAGCTGGCAGATC 650
339 nSerThrSerThrSerSerGlyGlyHisSerThrSerThrPheGlyThr 356
651 GGGCAATGCGCGGAGCCCTTCAACGGCCACTCAGATATCGTCAAAACA 700
356 hrSerGluThrProGluThrSerThrAspPheThrAlaThrSerThrSer 372
701 TCATCGCGCGCGAGAGAAATTTGTCGCGCAGGAGATGCGCTGCAGGAT 750
373 SerSerSerAspSerSerThrGlnSerSerAsnAla..... 384
751 ATAAGCGAAGGCTCAAAATTTGTCATGTCAGCGGCTTGGTCTGCTTTC 800
384 ..... 384
801 CACCGAAACACAGATGCGCGCATCAACGATTGCGCAGATATGGCGCAAC 850
385 .....GlnThrSerThrIle.....GluAsn 391
851 TCAAGACTATGCGCGCAGCAGCATCCGCGATTGGGCGAGTCCAAACCCC 900
392 GlySerThrThrAsnPheThrSerAlaProSerThrSerSerThrPr 408
901 ANTGGCCACAGGATAGAGCGGTCAGCAATATCTTTATGGCAGCCAT 950
408 oAla.....ThrProThrThrThrTyrAsnTrpProThrG 420
951 CCGCATCAAGGATTGGAGCTGTCGGGGAA.....AAT 985
420 lGlyThrThrTrpMetLeuProSerGlyGluLeValGlnAlaHisLys 436
986 ACGGCTGGGCGCATCACGGCACATC...CTGTCAAGCGGTCGACAGATG 1032
437 ThrPhePheSerSerPheAsnTyrPheGlnValLeuSerGluSerLeu1 453
```



```

322 erSerProHisThrProAsnProhe.....ArgProSerSerIleArgLys 336
935 ...TCITTTATGGAGCGCATCCCATCAAGGGA...TTGGAGCTGTCCGG 978
337 AspAlaLeuLeuGlnThrGlyProArgLeuGlyHisLeuGluCysLeuGI 353
979 GGAATATACGGCTTGGCGGCATCACGGACATCCCTGTCAAGCGGTCCGA 1028
353 yGlnProAlaAsnLeuArgThrSerGluArgSerProThrLysArgA 370
1029 GATGGCGCGCATGCTCAATTCGCGAAGGAAATCCG.....CCGTCA 1069
370 IgLeuProArgSerSerGluProAsnArgLeuProLysProLeuProGlu 386
1070 GCGACAATTTTGGCG...ATGCGGCAT.....ACGCC 1098
387 AlaThrLeuAlaProSerTyrArgHisArgSerTyrProPheLeuPr 403
1099 AAATACCGTCCCTTACCATTCCTCCGAAATATCCGTTCAACTTGGAGCA 1148
403 oAsnPro...ProAlaAlaLeuProSerIleAla.....TyrThrS 416
1149 CGCTTAGCGCAAGAAACATCCTCTCAACCGTCCCGCTCAACG 1198
416 er.SerArgGlyLysIleHisSerLeuProLysGlyAlaLeuProLy 432
1199 GCAAAATGTCAACTGGCAGACCAACGCCACCC 1232
432 sGlu.....GlyAlaProProPro 439
seq_name: SwissProt_40:YS8A_CAEEL

seq_documentation_block:
ID YS8A_CAEEL STANDARD; PRT; 796 AA.
AC O09625;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 84.3 kDa protein ZK945.10 in chromosome II.
GN ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48544; CAA88444.1; -
DR EMBL; Z48582; CAA88444.1; JOINED.
DR WormPep; ZK945.10; CE01732.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 30 POTENTIAL.
FT DOMAIN 273 546 SER/THR-RICH.
FT DOMAIN 656 752 SER/THR-RICH.
SQ SEQUENCE 796 AA; 84306 MW; 76DC5B03E6357A6A CRC64;

alignment_scores:
Quality: 124.00 Length: 395
Ratio: 0.649 Gaps: 17
Percent Similarity: 48.354 Percent Identity: 22.532
```



926 .....TCACAAATATCTTTATGGCAGCCATCCCATCAAG 961  
553 heTyrPheValGluLysAlaThrThrThrPheTyrAspSerThrSerVal 569  
962 GGATTGGAGCTGTCGGG  
570 AsnLeuThrLeuAsnSerGlyLeuGlyLeileGlyTyrGlnThrSerIl 586  
980 .GAAATACGCTGGCGGCATCAGCAGCATCTCTCAAGC.....G 1022  
586 eGluCysThrSerProThrSerSerAsnTyrValSerThrThrLysAspG 603  
1023 GTCCAGATGGCGCGCATCGCATCGCGAAG.....GGAATPCG 1063  
603 lYalaCysPheThrLysSerValSerMetProArgLeuGlyGlyThrTyr 619  
1064 CCCTCAGCAGCAATTG.....CCGATCGGCATAGCC 1098  
620 ProAlaSerThrPheValGlyProGlyAsnTyrThrPheArgAlaThrMe 636  
1099 AAATACCGCTCC.....CTTACCATCCCGAATATCCGTT 1136  
636 tThrAspAspLysValTyrThrTyrAlaAsnValTyrIleG 653  
1137 AAATTCGAGCGGTACGCGCAAGAAACATCA 1171  
653 InGluTyrSerThrThrIleGluSerGluSer 664  
seq\_name: SwissProt\_40:VGP3\_EBV

seq\_documentation\_block:

ID VGP3\_EBV STANDARD; PRT; 907 AA.

AC P03200; P03201;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Envelope glycoprotein GP340 (Membrane antigen) (MA) (Contains:

DE Glycoprotein GP220).

GN BLF1.

OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus.

OX NCBI\_TaxID=10377;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84270667; PubMed=6087149;

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,

RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,

RA Tuffnell P.S., Barrell B.G.;

RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.;"

RL Nature 310:207-211(1984).

CC -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN

CC B-CELLS.

CC -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL

CC ENVELOPE.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; V01555; CAA24854.1; -

DR PIR; A03762; Q0BE21.

DR PIR; A03763; Q0BE22.

DR PIR; S33008; S33008.

KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.

FT CARBOHYD 47 87 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 114 114 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 166 166 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 169 169 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 195 195 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 229 229 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 277 277 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 318 318 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 328 328 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 345 345 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 356 356 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 378 378 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 386 386 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 411 411 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 435 435 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 443 443 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 457 457 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 497 497 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 519 519 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 533 533 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 547 547 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 568 568 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 589 589 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 610 610 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 624 624 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 627 627 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 656 656 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 683 683 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 701 701 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 735 735 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 746 746 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 755 755 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 780 780 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 815 815 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 858 858 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 888 888 N-LINKED (GLCNAC...) (POTENTIAL).

FT VARSPLIC 502 698 MISSING (IN GP220).

SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC52C9 CRC64;

alignment\_scores:

Quality: 123.50 Length: 412

Ratio: 0.647 Gaps: 17

Percent Similarity: 46.359 Percent Identity: 22.330

alignment\_block:

US-09-303-518D-463 x VGP3\_EBV ..

Align seg 1/1 to: VGP3\_EBV from: 1 to: 907

110 TCGACCGTTCAGCATTTCCGACCGGGAATACCATTCGCGCAGC 159

469 SerThrAlaAspValThrSerProThrProAlaGlyThrThrSerGlyAl 485

160 AGGGGGGAGCTGTCNAGCGCAGCGGCATTCGATTGGGAACATACA 209

485 aSerProValThrProProProProProProProProProProProPro 496

210 AAGCCATCAGTGGCGCCACCTGATGATTAACACAGCGCGCGTTCAGGAA 259

497 .....AsnGlyThrGluSerLysAla 503

260 ATATCGCTACATGTCGCGTTTCCGATCAGC.....GGCAC 297

504 ProAspMetThrSerSerThrSerProValThrThrProThrProAsnAl 520

298 AAATTCATTCGCGCTTCGACACACCATTCATTCGATTCGATTCGACGA 347

520 aThrSerProThrProAlaValThrThrPro...ThrProAsnAlaThrS 536

348 AGCGGTAGTCCCGCTTACGCGATTTCAGCGCTTTACCGCATCCATTGGGACG 397

536 erProThrProAlaValThrThrProThrProAsnAlaThrSerProThr 552



## alignment\_scores:

Quality: 123.00 Length: 423  
 Ratio: 0.597 Gaps: 30  
 Percent Similarity: 48.700 Percent Identity: 24.823

## alignment\_block:

US-09-303-518D-463 x EGRI\_BRARE ..

Align seg 1/1 to: EGRI\_BRARE from: 1 to: 511

```

15 CAATAATCCCTTATCTGTCATACAGGAGTGTG.....CCTGC 55
||||| ||| ||||| |||||
90 GlnArgLeuProProLysSerThrGlyArgPheThrLeuGluProAl 106
56 CGATGTCATGCACAGCCTCAGATTGGCAACAGCATCCCTTTATCCGGCAG 105
||||| ||| ||||| |||||
106 aThrAsnCysSerAsn...SerLeuTrpAlaGlu...ProLeu..... 118
106 GTTCTCGACCGTC.....AGCA 122
||||| |||
119 .PheSerLeuValSerGlyLeuValGlyIleAsnProProAlaSerI 135
123 TTTTCGAACCGCAGCGGAATACACCTAT.....TCGGCAGAGGGGG 166
||| ||| ||||| ||||| ||||| |||||
135 LeProSerSerThrSerGlnAlaThrHisProSerSerSerThrSer 151
167 AGCTTGCNAGCGCAACGGCCATATCGGATTGGGAACATACAAAGCCAT 216
||||| ||||| ||||| ||||| |||||
152 SerIleProSerSerSerSerSer.....ThrSerSerAlaSerLe 166
217 CAGTTGGGCGACCTGATGATTCACAGCGCGCGTTGAAGAAATATCGG 266
||||| ||||| ||||| ||||| |||||
166 uSerCysSerValHisGlnSerGluProAsnProLleTyrSerAlaAlaP 183
267 CTACATGTCCGCT.....TTTCCGATCACGGGCACAAAT 301
||||| ||||| ||||| ||||| |||||
183 roThrTyrSerSerAlaSerProAspIlePheProGluSerGlyProAsn 199
302 .....TCCATTCCG.....CCTTCGACACCATGCC 327
200 PheSerThrValGlyThrSerLeuGlnTyrSerSerThrTyrPr 216
328 TCACATT.....CCGATTCTGACGAAGCGGTAGTCCCGTTGA 365
||| ||||| ||||| ||||| |||||
216 oSerAlaLysThrCysAsnProSerPheSerValProMetIlePro.... 231
366 CGGATTCACCTTTACCGCATCCATTGGACGGATACGAACACCATCCCG 415
||||| ||||| ||||| ||||| |||||
232 ..AspTyrLeuPheThrGlnGln...GlnSerGluIleSerLeuValPro 246
416 CGG.....ACGCTATGACGGCCACAGGGCGCGGCTAT 450
||||| ||||| ||||| ||||| |||||
247 ProAspGlnLysProLleGlnThrGlnAlaGlyGlnProAlaLeuTh 263
451 CCCGTCCTCCAAAGCGCGGAGGATATACAGTACGACATAAAAG... 497
||||| ||||| ||||| ||||| |||||
263 rProLeuHisThrIleLysAlaPheAlaThrGlnThr.GlySerGlnAsp 279
498 .....CGTTGCCCAAA 508
||| |||
280 LeuLysSerValTyrGlnSerGlnLeuLysProSerArgMetArgLy 296
509 ATATCCGCTCAACCTGACCGACAAACCGCAGCAG...CGGACACGGCTT 555
||||| ||||| ||||| ||||| |||||
296 sTyrProAsnArgProSerLysThrProProHisGluArgProTyrAlaC 313
556 GCCGACCGCTTTCCAAATGC...CGCGCTATGCTGACGCAAGGATAGG 602
||| ||| ||||| ||||| |||||
313 ys...ProValGluThrCysAspArgPhe.....SerArg 324
603 CGACGGATTCAACCGCGCCACCCCATACAGCCCGAGCTGGACAGTCGG 652
||||| ||||| ||||| ||||| |||||

```

```

325 SerAspGluLeuThrArgHis....IleArgIleHisThrGlyGln..... 338
653 GCAATGCCGCCGCAAGCCCTTCAACGGCACTGCAGATATCTCTCAAAACATC 702
||| ||||| ||||| |||||
339 .....LysProPheGln.....CysArgIleCysMetArgAsnP 350
703 ATCGCGCGCGGAGAGAAATTGTCGGCGCAGCGATCCGTCGAGGAT 752
||||| ||| ||||| ||||| |||||
350 heSerArgSerAspHisLeuThrThrHisIleArgThrHisThrGlyGlu 366
753 AAGCAAGGCTCAACATTTGCTCATGCACGCTTGGTGTCTGCTCTTCCA 802
||| ||| ||||| ||||| |||||
367 LysProPheAlaCysGluIleCysGlyArgLysPheAlaArgSerAspG1 383
803 CGAANAACACAGATGCCGCGCATCAACGATTTGGCAGATATGCGCAACTC 852
||||| ||| ||||| ||||| |||||
383 uArgLys.ArgHisThrLysIleHis.....MetArgGln... 394
853 AAAGACTATGCGCAGCAGCCATCCGCGATTGGCAGTCCCAAAACCCCAA 902
||||| ||| ||||| ||||| |||||
395 LysAspLysLysAlaGluLysGlyAlaThrAlaAlaValGlnSer..... 409
903 TCCGCGCACAAGGCATAGAACCGCTCAGCAATATC...TTTATGGCAGCCA 949
||| ||||| ||||| ||||| |||||
410 .....SerValSerAsnIleSerIleSerAlaSerS 420
950 TCCCATCAAGGATTTGGACCTCTCCGGGGAATAACGGCTTGGCGGC 999
||||| ||||| ||||| ||||| |||||
420 erProValSerSerTyrProSer.....Pro 428
1000 ATCAGCGCACATCTCTCAACGGCTCGCAGATGGCGCGATCGCATTTGCC 1049
||||| ||||| ||||| ||||| |||||
429 IleThrSerTyrProSerProValSerSerPhe..... 439
1050 GAAAGGAAATCCGCGCTCAGCGACAATTTTGGCGATCGCGCATACGCCA 1099
||| ||||| ||||| ||||| |||||
440 .....ProSerProValAsnSerCysTyrSerSerProValHisThrS 454
1100 AATACCGTCCCT 1113
||||| ||||| |||||
454 erTyrProSerPro 458
seq_name: SwissProt_40:A180_MOUSE
seq_documentation_block:
ID A180_MOUSE STANDARD; PRT; 901 AA.
AC Q61548; Q61547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clathrin coat assembly protein APl80 (Clathrin coat associated protein
DE APl80) (Phosphoprotein Fl-20) (91 kDa synaptosomal-associated
DE protein).
DE GN SNAP91.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92300439; PubMed=1607933;
RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
RT "Characterization of a novel synapse-specific protein. II. cdna
RT cloning and sequence analysis of the Fl-20 protein.";
RL J. Neurosci. 12:2144-2155(1992).
CC -!- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES WHICH
CC LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-
CC ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE
CC CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION
CC AND CONCENTRATION. BINDING OF APl80 TO CLATHRIN TRISKELIA INDUCES
CC THEIR ASSEMBLY INTO 60-70 NM COATS.
CC -!- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.

```



```
989 .....GTTGGCGGCATCAAGGCACATCTCTCAAGCGGTGCG 1027
    ||| |||:||||| |||:|||||
775 ysglyAspLeuGlnTtpAsnAlaGlyClulysLysLeuThrGlyGlyAla 791
    |||:||||| |||:|||||
1028 AGATGGCGGCATCGCATTCGCCGAAGGAATCCGCGGTACGACGAAT 1077
    :|||
792 AsnTrpGln..... 794
1078 TTTGGCGGATCGGCATACGCCAAATACCGTCCCTTACCATTCCCGAAA 1127
    |||:|||||
795 ..... 797
1128 TATCGGTCAAACTGTGAGCAGC...GTTACGCGCAAGAAACATCACCT 1174
    :|||:||||| |||:|||||
797 alThrProAlaThrTpsrAlaGlyValProGlnGlyThrValPro 813
1175 CTTCA....ACCGTGGCGGCTCAACGCGCAAAATGTCAAACCTGCCAGA 1220
    ||| :|||:||||| |||
814 ProThrSerValProProGlnGlyAlaGlyAla..... 824
1221 CCAACGCCACCCGAAGACAGCGGTACCGTTTGACGGGTAAAGGTTTCCGA 1270
    |||:||||| ||| ||| |||:|||||
825 .....ProSerValGlyClnProGlnGlyAlaGlyPheGlyMetProP 838
1271 ATTTGAGAAGCAGCTGAATATGATACGAAGCTCGATATCAAGAATTA 1320
838 ro..... 838
1321 TCGGGGCGGCTATACCT.....AAGGCTAAGCGTGTGTTTGTATGCGAA 1364
    ||||| |||:||||| |||:|||||
839 SerGlyThrGlyMetThrMetMetSerGlnGlnProValMetPheAlaG 855
1365 ACGG 1368
    :|||
855 nPro 856

seq_name: SwissProt_40:MGPC_MYCGE

seq_documentation_block:
ID MGPC_MYCGE STANDARD; PRT; 1052 AA.
AC P22747; Q49481; Q49482; Q49284; Q49385; Q49257;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MGPC protein precursor.
GN MGPC OR MG192.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=90060815; PubMed=2583522;
RA Inamine J.M., Loebel S., Collier A.M., Barile M.F., Hu P.-C.;
RT "Nucleotide sequence of the MgPa (mgp) operon of Mycoplasma
    genitalium and comparison to the Pl (mmp) operon of Mycoplasma
    pneumoniae.";
RL Gene 82:259-267 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
    Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
    Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
    Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
    Tomb J.-P., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
    Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403 (1995);
RN [3]
RP SEQUENCE OF 60-133; 260-370; 441-512 AND 964-1052.
```

```
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
    sequencing.";
RL J. Bacteriol. 175:7918-7930 (1993).
RN [4]
RP SEQUENCE OF 769-964 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=92051396; PubMed=1945886;
RA Peterson S.N., Schramm N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A random sequencing approach for placing markers on the physical map
    of Mycoplasma genitalium.";
RL Nucleic Acids Res. 19:6027-6031 (1991).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M31431; AAA25421.1; -
CC DR EMBL; U39698; AAC71411.1; -
CC DR EMBL; U02124; AAD12399.1; -
CC DR EMBL; U02161; AAD12443.1; -
CC DR EMBL; U34967; AAA88890.1; -
CC DR EMBL; U02157; AAD12439.1; -
CC DR EMBL; U34970; AAA88899.1; -
CC DR EMBL; X61525; CAB98130.1; -
CC DR EMBL; X61527; CAB98131.1; -
CC DR PIR; JQ0092; JQ0092.
CC DR TIGR; MG192; -.
CC KW Cell adhesion; Signal; Membrane; Complete proteome.
CC FT SIGNAL 1 25
CC FT CHAIN 26 1052
CC FT MGPC PROTEIN.
CC FT LDSSVQIAD -> MIKPLPLLS (IN REF. 3).
CC FT TM -> SV (IN REF. 3).
CC FT TNNGTGP -> SNOISSGT (IN REF. 3;
CC FT AAD12439).
CC FT N -> S (IN REF. 3; AAA88890).
CC FT SVSPRI -> MSHQGS (IN REF. 4).
CC FT SEQUENCE 1052 AA; 114360 MW; F6C6D3B3D7789145 CRC64;
SQ

alignment_scores:
    Quality: 123.00      Length: 686
    Ratio: 0.444        Gaps: 30
    Percent similarity: 40.379      Percent identity: 17.930

alignment_block:
US-09-303-518D-463 x MGPC_MYCGE ..
Align seg 1/1 to: MGPC_MYCGE from: 1 to: 1052
151 TTCGGCAGCAGGGGAGCTTGCNAGCGCAACGGCCATATC...GGATT 197
    :|||:||||| :|||:|||||
296 TyrAlaSerThrSerLysLeuAlaValThrAsnAsnHisIleValValMe 312
198 GGAAACATACAAAGCCATCAGTTGGCCACCTGATGATTCAACAGCGG 247
    :|||:||||| :|||:|||||
312 tGlyAsnSerPheLeuProSerMetTrpTrpValValGluArgSerA 329
248 CC..... 249
    ||
329 laGlnGluAsnAlaSerAsnLysProThrTrpPheAlaAsnThrAsnLeu 345
250 .....GTTGAAGGAATATCGGCTA 269
346 AspTrpGlyGluAspLysGlnLysGlnPheValGluAsnGlnLeuGlyTy 362
    :|||:||||| :|||:|||||
270 CATGTGCCCTTTTCGATCAGCGGCACAAATTCATTCGCCCC..... 312
```

[illegible][illegible]





857	heGluGlyPheSerAsnGlnAlaPheAlaThrLysLysGluGluTyr	873	seq_documentation_block:
652	GGCAATGCCCGCCGAAGCCTTCAAC.....GGCAC	680	ID SMF1_HUMAN STANDARD; PRT; 1902 AA.
874	ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluTrpGlyVa	890	AC Q1457; Q9UPZ1;
681	TGCAGATATCGTCAAAACATATCGCCGGCGGAGAGAAATGTCGGCG	730	DT 16-OCT-2001 (Rel. 40, Created)
890	lThrAsp.....PheGluMetAlaProGlnTyrValSerS	902	DT 16-OCT-2001 (Rel. 40, Last sequence update)
731	CAGCGCATGCCGNG.....CAGGATTAAGCGAAGCCTCAACAT	771	DT 16-OCT-2001 (Rel. 40, Last annotation update)
902	erThrAspGlySerPheLeuAspSerValIleGlnAsnGlyTyrAlaPhe	918	DE SWI/SNF-related, matrix-associated, actin-dependent regulator of
772	GCTGTCATGCAGCGCTGGTCTGCTTCCACCGAAGAAACAGATGCGCG	821	DE Chromatin subfamily F member 1 (SWI-SNF complex protein p270) (B120).
919	ThrAspArgTyrAspLeuGly...IleSerLysProAsnLysTyrGlyTh	934	GN SNARF1 OR Clorf4.
822	CATCAAGCATTTGCAGAT...ATGGCGCACTCAAGACTATGCGCGAG	868	OS Homo sapiens (Human).
934	rAlaAspLeuValLysAlaIleLysAlaLeuHisSerLysGlyIleL	951	OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
869	CAGCATTCGCGATTGGCGAGTCCAAACCCCAATGCCGACACAGGCATA	918	OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
951	ysValMetAlaAspTrpValProAspGln.....	960	OX NCBI_TaxID=9606;
919	GAAGCCGTGACCAATATCTTTATGGCAGCATCCCATCAAGGG...AT	965	RN [1]
961	MetTyrAlaLeuProGluLysGluValVa	970	SEQUENCE FROM N.A., AND MUTAGENESIS.
966	TGGAGCTGTCGG...GGAAATACGCTTGGGCGGCATCACGCGCATC	1012	RX MEDLINE=20221560; PubMed=10757798;
970	lThrAlaThrArgValAspLysTyrGlyThr.....P	981	RA Dallas P.B., Pacchione S., Willsker D., Bowrin V., Kobayashi R.,
1013	CTGTCAAGCGGTCCAGATG...GGCCGCGATTCGATTCGCGAAGGAAA	1059	RA Moran E.;
981	roValAlaGlySerGlnIleLysAsnThrLeuTyrValValAspGlyLys	997	RT "The human SWI-SNF complex protein p270 is an ARID family member with
1060	TCCGCGGTGACGACAAT.....TTGCGCGATGC	1088	RL non-sequence-specific DNA binding activity.";
998	SerSerGlyLysAspGlnGlnAlaLysTyrGlyGlyAlaPheLeuGlu	1014	RL Mol. Cell. Biol. 20:3137-3146(2000).
1089	GGCATACCCAAATACCGTCCCTTACCATCCCGAAATATCGTTCAA	1138	RN [2]
1014	uLeuGlnAlaLysTyrProGluLeuPheAlaArgLysGlnIle.....	1028	SEQUENCE OF 1-1175 FROM N.A.
1139	ACTTGGAGCAGCGTTACGGCAAGAAACATCATCTCTCAACCGTCCG	1188	RX MEDLINE=98094256; PubMed=9434167;
1029	SerThrGlyValPro	1033	RA Takeuchi T., Chen B.-K., Qiu Y., Sonobe H., Ohtsuki Y.;
1189	CCGTCA.....AACGG	1199	RT "Molecular cloning and expression of a novel human cDNA containing CAG
1034	MetAspProSerValLysIleLysGlnTrpSerAlaLysTyrPheAsnG	1050	repeats.";
1200	CAAAATGTC.....AACTGGCAGACCAACGCC	1228	Gene 204:71-77(1997).
1050	yThrAsnIleLeuGlyArgGlyAlaGlyTyrValLeuLysAspGlnAla	1067	[3]
1229	AC.....CCGAG...	1236	SEQUENCE OF 1-1132 FROM N.A.
1067	hrAsnThrTyrPheSerLeuValSerAspAsnThrPheLeuProLysSer	1083	Takeuchi T., Misaki A.;
1237	.....ACAGCGTACCGCT	1250	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
1084	LeuValAsnProAsnHisGlyThrSerSerValThrGlyLeuValPh	1100	CC -1- FUNCTION: BINDS DNA NON-SPECIFICALLY.
1251	TGACGGTAAAGGTTTCCGAATTTTGAGAAGCACGTGAATATGATACG	1299	CC -1- SUBUNIT: PART OF THE SWI-SNF COMPLEX.
1100	enaspGlyLysGlyTyr.....ValTyrTyrSerThr	1110	CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
			CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN, THYMUS, PROSTATE,
			CC LOWER LEVEL IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
			CC MUSCLE, KIDNEY, AND PANCREAS.
			CC -1- SIMILARITY: CONTAINS 1 ARID DOMAIN.
			CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1132
			CC ONWARD DUE TO A FRAMESHIFT.
			CC -----
			CC This SWISS-PROT entry is copyright. It is produced through a collaboration
			CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
			CC the European Bioinformatics Institute. There are no restrictions on its
			CC use by non-profit institutions as long as its content is in no way
			CC modified and this statement is not removed. Usage by and for commercial
			CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
			CC or send an email to license@isb-sib.ch).
			CC -----
			CC EMBL; AF265208; AAF75765.1; ALT_INIT.
			CC EMBL; AB001895; BAA23269.1; ALT_FRAME.
			CC EMBL; AB024075; BAA83073.1; ALT_SEQ.
			CC EMBL; AB024059; BAA83073.1; JOINED.
			CC EMBL; AB024060; BAA83073.1; JOINED.
			CC EMBL; AB024061; BAA83073.1; JOINED.
			CC EMBL; AB024062; BAA83073.1; JOINED.
			CC EMBL; AB024063; BAA83073.1; JOINED.
			CC EMBL; AB024064; BAA83073.1; JOINED.
			CC EMBL; AB024065; BAA83073.1; JOINED.
			CC EMBL; AB024066; BAA83073.1; JOINED.
			CC EMBL; AB024067; BAA83073.1; JOINED.
			CC EMBL; AB024068; BAA83073.1; JOINED.
			CC EMBL; AB024070; BAA83073.1; JOINED.
			CC EMBL; AB024071; BAA83073.1; JOINED.
			CC EMBL; AB024072; BAA83073.1; JOINED.
			CC EMBL; AB024073; BAA83073.1; JOINED.
			CC EMBL; AB024074; BAA83073.1; JOINED.
			CC MIM; 603024; -.
			CC InterPro: IPR001606; ARID.
			CC Pfam; PF01388; ARID; 1.



8 TTTCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTCCTGC GG 57  
||| ::::| | | | | | | | | | | | | | | | | | | | | |  
78 PheserProLeuHisProLeuValPheProGluHisHisSerGlnLeuAr 94  
58 ATGCATGCAC...ACGCCTCAGATTWGGCAAAGCATCCCTTATCGGCA 104  
| : : | | | | | | | | | | | | | | | | | | | | | |  
94 gHisValHisIglutThrGlnGlnValGlnThrCysPro..... 107  
105 GTTCTCGACCGTCAGCATTTTCGACCAGCGGAATAACCACTTATT.. 152  
108 ..GlvGlutLeuLvsLeuSerGlv.IteIglutLeuProProAlaPr 122

```

356 asLeuArgThrSerGluArgThrProSerThrLysArgLeuProA 373
1038 GA.....TCGATTGCCGAAGGAATCCGCGTCAGCAGCAATT 1078
373 rgProSerGluProAsnGlySerLeuProSerLeuProGluAlaThrLeu 389
1079 TTGCGG...ATGCGGCAT.....ACGCCAATATACCGC 1107
390 AlaProSerTyrArgHisArgArgSerHisProLeuLeuProAsnPro... 405
1108 TCCGCTTACCATTCGCCGAATATATCCGTTCAAACTTGAGCAGCGTTACGG 1157
406 ProAlaAlaLeuProProLeuAla.TyrThrSerGly.....Arg 418
1158 CAAGAAACATCATCTCTCCATCCGTCGCGCTCAAGCGCAAAATG 1207
419 GlyLysIleHisSerLeuProLysGlyAlaLeuProLysGlu..... 433
1208 TCAAACTGGCAGACCAACGCGCC 1232
434 .....GlyProProProProPro 439
seq_name: SwissProt_40:NCRL_HUMAN

```

## seq\_documentation\_block:

```

ID NCRL_HUMAN STANDARD; PRT: 2440 AA.
AC O75376; Q90PV5; Q90Q18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Nuclear receptor co-repressor 1 (N-Cor1) (N-Cor).
GN NCRL OR KIAA1047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98393736; PubMed=9724795;
RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;
RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
RT transcription by interaction with the human N-Cor/msln3/HDAC1
RT complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
RN [2]
RP SEQUENCE OF 782-2440 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
RN [3]
RP SEQUENCE OF 974-2440 FROM N.A.
RX MEDLINE=99375328; PubMed=1044336;
RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,
RA Horwitz K.B., Lupski J.R., Seo H.;
RT "Localization of the human nuclear receptor co-repressor (hn-Cor) gene
RT between the CMT1A and the SMS critical regions of chromosome
RL 17p11.2.";
RL Genomics 59:339-341(1999).
CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

```

```

-1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
DOMAINS (ID1 AND ID2).
-1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
SPECIFICITY.
-1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
-1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
-1- SIMILARITY: CONTAINS 2 CORNR BOX.
-1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF044209; AAC33550.1; -
EMBL; AB028970; BAA82999.1; -
EMBL; AB019524; BAA75814.1; -
MTM; 600849; -
InterPro: IPR001005; Myb_DNA_bind.
Pfam: PF00249; myb_DNA-binding; 2.
SMART; SM00395; SANT; 2.
PROSITE; PS50090; MYB_3; 1.
Nuclear protein; Transcription regulation; DNA-binding; Repressor;
Coiled coil.
KW DOMAIN 174 216 COILED COIL (POTENTIAL).
FT DOMAIN 254 312 INTERACTION WITH SIN3A/B.
FT DOMAIN 299 328 COILED COIL (POTENTIAL).
FT DNA_BIND 437 482 SANT-A (POTENTIAL).
FT MYB 620 670 MYB.
FT DOMAIN 501 557 COILED COIL (POTENTIAL).
FT DOMAIN 607 617 PRO-RICH.
FT DOMAIN 988 1816 INTERACTION WITH ETO.
FT DOMAIN 2055 2059 CORNR BOX OF ID1.
FT DOMAIN 2263 2267 CORNR BOX OF ID2.
FT DOMAIN 58 64 POLY-GLN.
FT DOMAIN 593 603 POLY-ALA.
FT DOMAIN 1032 1035 POLY-PRO.
FT DOMAIN 1707 1712 POLY-ALA.
FT DOMAIN 1952 1963 POLY-SER.
FT CONFLICT 1014 1014 L -> V (IN REF. 2).
FT CONFLICT 1508 1509 PP -> SS (IN REF. 2).
FT CONFLICT 1561 1561 W -> R (IN REF. 2).
FT CONFLICT 1567 1567 Q -> H (IN REF. 2).
SQ SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

```

## alignment\_scores:

```

Quality: 121.50 Length: 646
Ratio: 0.428 Gaps: 31
Percent similarity: 43.963 Percent identity: 19.969

```

## alignment\_block:

```

US-09-303-518D-463 x NCRL_HUMAN
Align seg 1/1 to: NCRL_HUMAN from: 1 to: 2440
12 CCAGAAATATCCCTTATCTGTCATCTGCGCAGCTG.....CCTGC 55
||| :|||:||||| ||| :|||:||||| |||
924 ProLeuAspLeuProGlnLeuGlnHisArgAlaAlaValLeuProPme 940
56 CGATGATGCACACGCTCATGTTTGCGCAACGATCCCTTTATCCGCGAG 105
:|||||:||||| :|||:
940 tValSerCysThrProCysAsnIle.....ProIleG 951

```

106 GTTCTCGACGTCAGCATTT.....CGAACCGGA 134  
1176 ylleProThrGluAlaLeuValLysGlySerIleSerArgMetProIleG 1193  
942 GGCAGCCATCCCATCAAGGATTCGAGCTGTCGGGGGAAAAATACGGCT 991  
1193 luaspSerSerProGluLysGlyArgGluAla..... 1204  
992 TGGCGGCATCACGGCACATCTCTCAAGGGTTCGACATGGGCGCGCATC 1041  
1205 ...AlaSerLysGlyHisValIleThrGluGlyLys..... 1215  
1042 GCATTGCCGAAGGAAATCCCGCTCAGCGACAAATTTCCCGATGCGGC 1091  
1216 .....SerGlyHisIleLeuSerTyrAspA 1224  
1092 ATACGCCCAATACCCGT.....CCCTTACCATTCGCCGAA 1126  
1224 snlleLysAsnAlaArgGluGlyThrArgSerProArgThrAlaHisGlu 1240  
1127 ATATCCGTT...CAAACTTGGAGCAGCTTACGGC.....AAGAAAA 1166  
1241 ileSerLeuLysArgSerTyrGluSerValGluGlyAsnIleLysGlnG 1257  
1167 CATCACCTCTCAACCGTCCG.....CCGTCAACCGCAAAAAATG 1207  
1257 yMetSerMetArgGluSerProValSerAlaProLeuGluGlyLeuIleC 1274  
1208 TCAAACTGCGACAGCAACGCCACCG.....AAGACAGCGGTA 1245  
1274 ysArgAlaLeuProArgGlySerProHisSerAspLeuLysGluArgThr 1290  
1246 CGCTTTGACGGT.....AAGGGTTTCGG.....AA 1271  
1291 valLeuSerGlySerIleMetGlnGlyThrProArgAlaThrThrGluse 1307  
1272 TTTTGAAGACACGTGAATATGATACGAAGCTCGATATTCAA..... 1314  
1307 rPheGluaspGlyLeuLysTyrProLysGlnIleLysArgGluSerProp 1324  
1315 ....GAATTATCGGGGGCGGTATACCTAAGCTTAAAGCTGTGTGTAT 1359  
1324 roIleArgAlaPheGluGlyAlaIleThrLysGlyLysProTyrAspGly 1340  
1360 GCGAAACCGAGATGGAGGTTCATAGGAGCTTAATAATTGACAACTCG 1409  
1341 ileThrThrIleLysGluMetGlyArgSerIleHisGluIleProArgG 1357  
1410 TGAGCAGGTGGAGAAAAATGTTCAAGAAACAGAGAGGAGTCAGAGTA 1459  
1357 nAspIleLeuThrGlnGluSerArgLysThrProGluValValGlnSerT 1374  
1460 GTCAG.....TTTAAAGCCCATCGCAACGAGAAATGG 1491  
1374 hrArgProIleIleGluGlySerIleSerGlnGlyThrProIleLysPhe 1390  
1492 GAAATATAAACAGGGTTAGAT...TTTATCATTTTATAGTGGTGCATAT 1538  
1391 AspAsnAsnSerGlyGlnSerAlaIleLysHis..... 1401  
1539 CAATAAGAGGACACAGTACAGGAGGCGCATAGCTAACCCCGTGT..... 1584  
1402 .AsnValLysSerLeuIleThrGlyProSerLysLeuSerArgGlyMetP 1418  
1585 ..GATGTACGGGTGATACAAACAAACCTCGGCACCTGATAAACATGGGGT 1632  
1418 roProLeuGluIleValProGluAsnIleLysValValGluArgGlyLys 1434  
1633 TATCAAGCGCAGCTGGAAATTAATAAACCGCTGATGGAAGTTGGAGGTGAA 1682  
1435 TyrGlu.....As 1437  
1683 AACGAAAAAGGTGGGAAAGTGATGACCAAGCACACC 1719

106 GTTCTCGACGTCAGCATTT.....CGAACCGGA 134  
951 lyThrProValSerGlyTyrAlaLeuTyrGlnArgHisIleLysAlaMet 967  
135 CGGAAATACCACTTATTCGGCAGCAGGGGAGGTTCGCCNAGCGCAACG 184  
968 HisGluSerAlaLeuLeuGluGln.....ArgGlnAr 979  
185 GCATATCGGATGGGAACATACAAAGCCATCAGTTCGGCCACCTGATG 234  
979 gGlnGluGlnIleAspLeuGluCysArgSerThrSerProCysGlyT 996  
235 ATTCACAGCGCGCTTGAAGGAATATCATCTGCTACATTTGCTCGCTTTTC 284  
996 hrSer...LysSerProAsnArgGluTrpGluValLeuGlnProAlaProH 1012  
285 CGATCAGCGGCACAAAT.....TCCATTCGCCCTTCGACACCATG 325  
1012 isGlnLeuIleThrAsnLeuProGluGlyValArgLeuProThrThrArg 1028  
326 CTTACATTCGATTCGAGGAGCGGTAGTCCCGTTGACGGATTCAGC 375  
1029 Pro.....ThrArgProProProProLeuIleProSerSe 1040  
376 CTTTACCGCATCCATT.....GGGACGG 398  
1040 rLysThrThrValAlaSerGluLysProSerPheIleMetGlyGlySerI 1057  
399 ATAGAACACCATCCCGCCGACGCTATGACGGGCGACAGGGCGGGCT 448  
1057 leSerGlnGlyThrProGlyThrTyrLeuThrSerHisAsnGlnAlaSer 1073  
449 ATCCCGCTCCCAAGGCGGAGGATATATACAGTACGACATATAAAGGC 498  
1074 .....TyrThrGlnGluThr..... 1078  
499 GTTCCCAAAATATCCCGCTCAACCTGACCGACACCGCACCGCGACA 548  
1079 ....ProLys.....ProSerValGlySerIleSerL 1088  
549 ACGCTTTCGCGACGCTTCCACATGCGCGGTATGCTCAGCAGGAG 598  
1088 euGlyLeuProArg.....GlnGlnGlu 1095  
599 TAGCGCGGATTCAAACGCGCCACCGATACAGCCCGCGAGCTGGACAGA 648  
1096 SerAlaLysSerAlaThrLeuProTyrIleLysGlnGluGluPheSerPr 1112  
649 TCGGCAATGCGCGCGAGCT..... 670  
1112 oArgSerGlnAsnSerGlnProGluGlyLeuLeuValArgAlaGlnHisG 1129  
671 .....TCAACGGCAGTGCAGATATCGTCAAAACA 700  
1129 luGlyValValArgGlyThrAlaGlyAlaIleGlnGluGlySerIleThr 1145  
701 TCATCGCGCGCGCAGGAAATTCGCGCGCAGCGGATGCGTGCAGGGT 750  
1146 ArgGlyThrProThrSerLysIleSerValGluSerIlePro..... 1159  
751 ATAAGCGAAGGCTCAAAATGCTGTCTATGACGCGCTTGGGTCTGCTTTC 800  
1159 ..... 1159  
801 CACCGAAAAACAAGATGGCGCGCATCAACGATTGGCAGATATGGCGAAC 850  
1160 .....SerLeuArgGlySerIleThrGlnGly 1168  
851 TCAAGACTATGCCGACGACCCATCCCGGATGGCGCAGTCCAAACCC 900  
1169 ThrProAlaLeuProGln.....ThrGl 1176

seq\_name: SwissProt\_40:PSC\_DROME  
1437 pValLysalacGlyGluThrValArgSerArgHisThr 1449

Align seg 1/1 to: PSC\_DROME from: 1 to: 1603

```
seq_documentation_block:
ID   PSC_DROME          STANDARD;          PRT; 1603 AA.
AC   P35820;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Posterior sex combs protein.
GN   PSC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92018190; PubMed=1833647;
RA   Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT   "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RL   Nature 353:351-353(1991).
CC   FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC   EXPRESSION PATTERNS OF THE HOMEOTIC SELECTOR GENES OF THE
CC   ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
CC   THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC   SUBCELLULAR LOCATION: Nuclear (Probable).
CC   SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X59275; CAA41965.1; -
DR   PIR; S17983; S17983.
DR   FlyBase; FBgn0005624; Psc.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; ZF-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT   DOMAIN 47 53 POLY-THR.
FT   DOMAIN 83 88 POLY-THR.
FT   DOMAIN 91 98 POLY-THR.
FT   DOMAIN 145 152 POLY-THR.
FT   DOMAIN 184 202 POLY-SER.
FT   ZN_FING 265 304 RING-TYPE.
FT   DOMAIN 642 651 POLY-SER.
FT   DOMAIN 1066 1069 POLY-GLY.
FT   DOMAIN 1185 1189 POLY-PRO.
FT   DOMAIN 1214 1217 POLY-PRO.
FT   DOMAIN 1391 1396 POLY-PRO.
FT   DOMAIN 1458 1461 POLY-ALA.
FT   DOMAIN 1517 1520 POLY-GLY.
SQ   SEQUENCE 1603 AA; 16999 MW; 77024F4097736473 CRC64;

alignment_scores:
Quality: 121.00 Length: 361
Ratio: 0.761 Gaps: 15
Percent Similarity: 44.044 Percent Identity: 22.161

alignment_block:
US-09-303-518D-463 x PSC_DROME
275 TCGGCTTTTCGATCACGGGCACAAATTCATTCGCCCTTCGACAAACCAT 324
1128 SerSerProAsnAsnGlyGlnHisProLysHisLysSerProValAs 1144
325 GCCTCACATTCGATTCGACGAACCGGAGTAGTCGCGTTCGAGGATTCAG 374
1144 nAsnTyrlleGluilleValLysLeuProAspGlnPro.....GlnAspG 1159
375 CCTTTACCGCATCCATTGGGACGATAGACACACCATCCCGCGCAGCGCT 424
1159 InValGlnAlaAlaLysGluAlaGlnLysArgGlnSerProProAlaAla 1175
425 ATGACGGGCGACAGCGCGCGGCGGTATCCCG..... 454
1176 ValProGlyHisLeuAlaAlaLysLeuProProProProSerLysAl 1192
455 .CTCCCAAGGCGCGAGGATATATACAGCTACGACATAAAGCGGTGC 503
1192 aileProSerProGlnHisLeuValSerArgMetThrProProGlnLeuP 1209
504 CCAAAATATCCGCTCA.....ACCTGA 526
1209 toLysValAlaThrProProProProSerProArgValilleThrPro 1225
527 CCGACACCGCGACGACCGGACAGCGGCTGCGGACGTTTCCACAATGCC 576
1226 ProLysThrSerProProAlaAlaAlaLysValThrProLeuLysPr 1242
577 GCGCTATGCTGACGCAAGGAGTAGGCGACGAGTCAACACGCGCCACCG 626
1242 oValLeu..... 1244
627 ATACAGCCCGGACGATGACATCGGCAATGCG.....CGAGCCCTCA 673
1245 .....ThrProThrGlnValAsnLysLysThrProSerProGluLysArg 1259
674 ACGGCACTGCGATATCGTCAAAACATCATCGCGCGGAGGAGAAATT 722
1260 ThrAlaAlaGlnMetGlySerHisSerProThrAlaSerGluAsnLysSe 1276
723 .....GTCCGCGCAGCGATGCGTGCAGGAGGTATAAGC 757
1276 rProLysGlyGlyAlaAlaGlyValAlaAsnSerThrGlyGlyThrGlnA 1293
758 AAGGCTCAACATTCGCTGTCATCGACGCGTGGTCTGCTTCCACCGAA 807
1293 snGlyAspProAlaAla.....LysLysPheArg 1302
808 AACAGATGGCGCATCAACGATTTGGCAGATATGGCGCACTCAAGA 857
1303 ProfileLeuProArgGlnAsnGlyMetProGluLeuAlaProLysLeuPr 1319
858 CTATGCGCGCAGCGCATCGCGGATGGGAGTCCCAAAACCCCAATGCCG 907
1319 oThrLeuAlaProPheValGlyPheAsnProLeuGlnAsnProAlaAlaG 1336
908 CACAAGCATAGAAAGCGTCAGCAATATCTTTATGGCAGCCATCCCATC 957
1336 lyLysLysVal.ProProSer.....LysLysSerProAs 1347
958 AAAGGATGGAGCTCTCCGGGAAATACGGCTTGGGCGGCATCACGGC 1007
1347 nAlaGly.....AlaAlaAla 1353
1008 ACATCTGTCAAGCGTTCGAGATGGCGCGCATTCGCGCAAGGGA 1057
1353 is.....GlnSerGlyGln.....GlnLysLeuValAsnGlyGly 1364
1058 AATCCGCGCTCAGCGACAATTTTGGCGGATGGCGCATACGCAATACCG 1107
1365 GlnProGlnSerAla..... 1369
```





```

1445 GlnThrAsnSerThrValProProSerAla...ProProProThrThrAl 1460
445 GCCTATCCCGCTCCCAAGGCG.....CGAGGGATAT 476
1460 aAlaThrProLeuProThrSerPheProThrLeuSerPheGlySerLeuL 1477
477 ATACAGCTACACATATAAGCGCTTGCCCAAAATATCCGCTCAACCTGA 526
1477 euSerSerAlaThrProSerLeuProMetSerAlaGlyArgSer... 1492
527 CCGCAACCGCAGCAGCGGACACGCGCTGCGG..... 559
1493 ...ThrGluGluAlaThrSerSerAlaLeuProGluLysProGlyAspSe 1508
560 ...ACGTTTCCCAATCGCGCGCTATGCTGACGCAAGGAGTAGCGGA 605
1508 rGluValSerAlaSerAlaAlaSerLeuGluGluGlnGlnSerAlaG 1525
606 CGGATTCAACCGCCACCGATACACGCCCGGAGCTGGAGATCGGGCA 655
1525 lnLeuProGlnAlaProProGlnThrSerAspSerValLysLysGluPro 1541
656 ATGCGCGCGGAGCGCTTCAACGGCAGCTGATCGTCAAAAACATCATC 705
1542 ValLeuAlaGlnProAlaValSer...AsnSerGlyThrAlaAlaSerSe 1557
706 GCGCGCGCAGAGAAATTCGCGCGAGCGGATGCGCTGCGAGGGTATAAG 755
1557 rThrSerLeuValAlaLeuSerAlaGluAlaThrPro..... 1569
756 CGAAGGCTCAACATTCCTGCTGATGACGCGCTGGTCTGCTTCCACCG 805
1570 .....AlaThrThrGlyValProAspAlaArgThrGluAlaValPro 1584
806 AAACAAGATGCGCGCATCAACGATTTGCGAGATGCGCAACTCAAA 855
1585 .....AlaSerSerPh 1588
856 GACTATGCGCGCAGCAG..... 871
1588 eSerValProGlyGlnThrAlaValThrAlaAlaLeuSerSerAlaG 1605
872 ..CCATCGCGGATGGGAGTCCAAACCCCAATGCCGCAAGGATAG 919
1605 lyProValAlaValGluThrSerSerThrProLe..... 1616
920 AACCCCTCAGCATATCTTATGGCAGCATCCCATCAAGGGATGGA 969
1617 AlaSerThrThrSerThrSerThrSerThrSerThrSerThrSerThr 1633
970 GCTGTCCGGGAAATACGCTTGGCG.....GCATCAGCGCAC 1010
1633 aAlaPheGlyThrValThrSerGlySerSerValPheAlaGlnProPro 1650
1011 TCCTGTCAAGCGGTCCAGATGGCGCGCATGTCGCGCAAGGGAAT 1060
1650 laAlaSerSerSerAlaPheAsnGlnLeuThrAsnAsnThrAlaThr 1666
1061 CGCGCTCAGCGACATTTTCCGATGCGGATACGCCAAATACCCGTC 1110
1667 AlaProSerAlaThrProValPheGlyGlnValAlaAlaSerThrAlaPr 1683
1111 C.....CTTACCA 1118
1683 oSerLeuPheGlyGlnGlnThrGlySerThrAlaSerThrAlaAlaAla 1700
1119 TTCGCAATATCCGTTCAACTTGAGCAGCGTTACGCGCAAGAAACA 1168
1700 hrProGlnValSerSerGlyPheSerSerProAlaPheGlyThrThr 1716
1169 TCACCT 1174
1717 AlaPro 1718

```

seq\_name: SwissProt\_40:ALP\_CEPAC

seq\_documentation\_block:

ID ALP\_CEPAC STANDARD: PRT; 402 AA.

AC P29118;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).

GN ALP.

OS Cephalosporium acremonium (Acremonium chrysogenum).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acremonium.

OX NCBI\_TaxID=5044;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91299283; PubMed=1368696;

RA Isogai T., Fukagawa M., Kojo H., Kohsaka M., Aoki H., Imanaka H.;

RT "Cloning and nucleotide sequences of the complementary and genomic

DNAs for the alkaline proteinase from Acremonium chrysogenum.";

RL Agric. Biol. Chem. 55:471-477(1991).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

CC SUBTILASE FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC EMBL; D00923; BAA00765.1;

DR PIR; J00332; J00332.

DR HSP; P06873; 2PRK.

DR InterPro; IPR000209; Peptidase\_S8.

DR Pfam; PF00082; Peptidase\_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE-ASP; 1.

DR PROSITE; PS00137; SUBTILASE-HIS; 1.

DR PROSITE; PS00138; SUBTILASE-SER; 1.

KW Hydrolase; Serine protease; zymogen; signal.

FT SIGNAL 1 20 POTENTIAL.

FT PROPEP 21 120 POTENTIAL.

FT CHAIN 121 402 ALKALINE PROTEINASE.

FT ACT\_SITE 160 160 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 191 191 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 347 347 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT SEQUENCE 402 AA; 42099 MW; 8D030CCD42D918E1 CRC64;

SQ

# alignment\_scores:

Quality: 120.00 Length: 324  
 Ratio: 0.745 Gaps: 15  
 Percent Similarity: 49.691 Percent Identity: 22.222

# alignment\_block:

US-09-303-518D-463 x ALP\_CEPAC ..

Align seg 1/1 to: ALP\_CEPAC from: 1 to: 402

```

76 GATTGGCAAAAGATCCCTTTATCCGCGAGTTCTCGACCGTCAGCATTT 125
   ::::: ||:::| |:::| |:::| |:::| |:::| |:::| |:::|
90 GluIleLysAsnAsnProAspValLeuGluVal..... 100

126 CGAACCGCGCGGAAATACACCATATTCGCCACACGCGGGGAG..... 168
   ||| ||| ::::| |:::| |:::| |:::| |:::| |:::| |:::|
101 .GluGluAspGlnIleThrPheHisLeuPheAspGluGlnAspGluGly 117
169 .....CTGCCNAGCGCAAGCCATATCGATGTTGGGAAC 204
   ||:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
117 heSerThrAlaAlaLeuValThrGlnAsnGlyAlaTrpGlyLeuGlyThr 133

```









Percent Similarity: 45.198 Percent Identity: 21.610

alignment\_block:

US-09-303-518D-463 x CAIE\_CHECK ..

Align seg 1/1 to: CAIE\_CHECK from: 1 to: 1888

```

83 CAACGATCCCTTTATCCGGCAGGTTCTCACCGTCAGCATTTGCAACCC 132
   ::::: |||||
667 GluGluValValLeuSerGlyAspAlaAspSerTyrValIleGluGlyLe 683
   ::::: |||||
133 GACGGGAATACCACTATTCGGCAGCAGGGGGAGTTCGCCNAGCGCAA 182
   ||||| |||||
683 uLeuProAsnThrGluTyrValIleGluGlyValIleGluGlyValIle 694
   ||||| |||||
183 CGGCATATCGGATTTGGGAACATACAAAGCCATCAGTTGGGCCACCTGA 232
   ||||| |||||
694 euAlaValPheAspAspGluThrGluSerGluValIleGluGlyValIle 705
   ||||| |||||
233 TGATTCAACAGCGCGCTTGAAGGAATATCGGCTACATTG..... 274
   ::::: |||||
706 .....ValAlaValLeuGlyAlaThrIleValGlyTh 716
   ::::: |||||
275 .TCGCTTTTCGATCACGGGCACAAATTCATTCGCCCTTCGACAAACCA 323
   ::|||::||| |||||
716 rThrAlaIleProThrValThrThrThrThrThrThrThrAlaThrT 733
   ::|||::||| |||||
324 TGCCTCACATTCGATTCGACGAAGCGGTAGTCCCG..... 361
   ||||| |||||
733 hrPro.....LysProThrIleAlaValPheArgThr 743
   ||||| |||||
362 .....TTGACGGATTCAGCCTTTACCGCATCCATTGGG... 394
   ::|||::||| |||||
744 GlyValArgAsnLeuValIleAspAspGluThrThrSerSerLeuArgVa 760
   ::|||::||| |||||
395 .....ACGGATAGCAACACC.....ATCCCGCCGA 419
   ::|||::||| |||||
760 lValTrpAspIleSerAspHisAsnAlaGlnGlnPheArgValThrTyrL 777
   ::|||::||| |||||
420 CGGCTATGACGGGCACAGGCGG...GGGGTATCCGCTCCCAAGGCG 466
   ::|||::||| |||||
777 euThrAlaLysGlyAspArgAlaGluGluAlaIleMetValPro.GlyAr 793
   ::|||::||| |||||
467 CGAGGGATATACAGCTACGACATAAAAGCGTTGCCAAATATCCGC 516
   ::|||::||| |||||
793 gGlnAsnThrLeuLeuLeuGlnProLeuLeuProAspThrGluTyr.Lys 809
   ::|||::||| |||||
517 CTCAACCTGACCGACACCGCAGCACCAGCA..... 549
   ::|||::||| |||||
810 ValThrIleThrProIleTyrAlaAspGlyGluGlyValSerValSerAl 826
   ::|||::||| |||||
550 .....CGGCTGCGG 559
   ::|||::||| |||||
826 aProGlyLysThrLeuProLeuSerAlaProArgAsnLeuArgValSerA 843
   ::|||::||| |||||
560 ACCGTTTCACAAAT.....GCCGGCGCTATGCTG 588
   ::|||::||| |||||
843 sPgluTyrTrpAsnArgLeuArgIleSerTrpAspAlaProProSerPro 859
   ::|||::||| |||||
589 ACACAAGGATAGGCGCAGGATTCAAACCGCCACCCGATACAGCCCGCA 638
   ::|||::||| |||||
860 ThrMetGlyTyrArgIleValTyrLys.....Se 869
   ::|||::||| |||||
639 GCTGGACAGATCGGCAATCGCGCAGACCGCTTCAACGCG.....A 679
   ::|||::||| |||||
869 rIleAsnValProGlyProAlaLeuGluThrPheValGlyAspAspIleA 886
   ::|||::||| |||||
680 CTGCAGATATCTCAAAACATCATCGGCGCG..... 711
   ::|||::||| |||||
886 snThrIleLeuLeuLeuAsnLeuPheSerGlyThrGluTyrSerValLys 902
   ::|||::||| |||||
712 .....GCAGGAGAAATTCGCGCAGCGGATCGGTCAGGGTATAAG 755
   ||||| ::|||::||| |||||

```

```

903 ValPheAlaSerTyrSerThrGlyPheSerAspAlaLeuThrGlyValAl 919
756 CGAAGGCTCAAAACATTGCTGTCAATGCACGGCTTGGGTCTGCTTTCCACCG 805
   ::::: |||||
919 aLys.....ThrLeuTyr...LeuGlyValThrAsnLeuA 930
806 AAACAAGATGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAA 855
   ::::: |||||
930 sPThrTyrGlnValArgMetThrSerLeuCysAlaGlnTrpGlnLeuHis 946
   ::|||::||| |||||
856 GACTATGCGCGCAGCA.....GCCATCCGCGATTGGCG 887
   ::|||::||| |||||
947 ArgHisAlaThrAlaTyrArgValValIleGluSerLeuValAspGlyLy 963
888 AGTCAAAACCCCAATCGCGCAAGGCATAGAACCGCTCAGCAATATCT 937
   ||||| |||||
963 sLysGlnGluValAsnLeuGlyGlyValPro.....ArgHisCysP 978
938 TTATGCGACCATCCCATCAAAAGGNTTGGAGCTGTCCGGGGAATAATAC 987
   ||||| |||||
978 hePheGluLeuMetPro.....GlyThrGluTyr 987
988 GCGTTGGCGGCATCACGCCACATCTGTCAAGCGGTGCGAGATGGCGC 1037
   ||||| |||||
988 .....LysIleSerValHis.....AlaGlnLeuGlnI 997
1038 GATCGCATTTCCGAAAGGAAATCCGCGTCAGCGACAAATTTTGGCGATG 1087
   ||||| |||||
997 uIleGluGlyPro...AlaValSerIleMetGlu.ThrThrLeuProPhe 1012
1088 CGGCATACCCAAATACCGTCCCTTACCATTCGCGAAATATCCGTTC 1137
   ||||| |||||
1013 ProThrGlnProProThrSerProSerThr..... 1022
1138 AACTTGGACAGCGTTACGGCAAGAAACATCACCTCTCTCA. ACCTGTC 1186
1023 .....ThrLeuProProProThrIleP 1030
1187 CGCGCTCAAAACGCAAAATGTCAAACTGGCAGACCAACGCCACCGGAAG 1236
   ||||| |||||
1030 roProAla.....LysGluValCysLysAla.....AlaLys 1040
1237 ACAGGCGTACCGTTT.....GACGGTAA.....GGTTTCCGAA 1271
   ::|||::||| |||||
1041 AlaAspLeuValPheLeuValAspGlySerTrpSerIleGlyAspAspAs 1057
1272 TTTTGAGAACACGCTGAAA.....TATGATACCAAGCTCGATATTCAG 1315
   ||||| |||||
1057 nPheAsnLysIleIleSerPheLeuTyrSerThrValGlyAlaLeuAspL 1074
1316 AATTATCGGGGGCGGTATACCTAAGGCTAAGCTGTGTTTGTATGCGAAA 1365
   ::|||::||| |||||
1074 ysIleGlyProAspGlyThrGlnValAlaIleIleGlnPheSerAspAsp 1090
1366 CCGAGATGGGAGGTGTATAGACGCTTAATAATTGACAACTCTGAGCA 1415
   ||||| |||||
1091 ProArgThrGluPhe.....LysLeuAsnAlaTyrLysThrLysGluTh 1105
1416 GGTGGGAAAAATCTTCAGGAACGAGAGAGAGAGTACAGATAGTACAGT 1465
   ::|||::||| |||||
1105 rLeuLeuGluAlaIleGlnIle..... 1113
1466 TTAAGCCCATGCGCAACGAGAATGGGAAATAAACAAGGTTAGATTTT 1515
   ||||| |||||
1114 .....AlaTyrLysGlyGlyAsnThrLysThrGlyLysAlaIle 1126
1516 AATCAT.....TTTATAGGTGGTGCATATCAATCAATGAAGAAG 1550
   ::|||::||| |||||
1127 LysHisAlaArgGluValLeuPheThrGlyGluAlaGlyMetArgLysGI 1143
1551 C.....ACAGTAACAGGAGGCGCATAGTCTAACCCGTG 1582
   ||||| ::|||::||| |||||
1143 yIleProLysValLeuValValIleThrAspGlyArgSer.....GlnA 1158

```





```
337 pSerAlaLysIleGlyAlaAspValLeuAsnMetSerLeuGlySerAsps 354
611 TCAACGGCCACCCGATACAGCCGAGCTGGACAGATCGGGCAATGCC 660
354 erGlyAsnGlnThrLeuGluAspProGluLeuAlaValGlnAsnAla 370
661 GCCAAGCCCTTCAACGGCCTCAGATATCGTCAAAACATCATCGGCC 710
371 AsnGlu.....SerGlyThrAlaAlaValIle.....Se 380
711 GCCAGGAGAAATGTCCGCGCAGCGCATCGCTGCAGGGTATA..... 753
380 rAlaGlyAsnSerGlyThrSerGlySerAlaThrGluGlyValAsnLysA 397
754 ..... 768
397 spTyrtYrGlyLeuGlnAspAsnGluMetValGlyThrProGlyThrSer 413
757 GAAGGCTCAAC..... 768
414 ArgGlyAlaThrThrValAlaSerAlaGluAsnThrAspValIleThrGl 430
769 .....ATTGCTGTCATGCACGGCTTGGGTCTG..... 795
430 nAlaValIleThrAspGlyThrGlyLeuGlnLeuGlyProGlyThrI 447
796 .....CTTCCACC.....GAAACACAGATG 816
447 leGlnLeuSerSerAsnAspPheThrGlySerPheAspGlnLysLysPhe 463
817 CGCGCATCAACGATTGGCA.....GATATGGCGCACTCAAGA 857
464 TyrValValLysAspAlaSerGlyAsnLeuSerLysGlyAlaLeuAlaAs 480
858 CTATGCGCGCAGCAGCATCCGCGATTGGCGATC..... 891
480 pTyThrAlaAspAlaLysGlyLysIleAlaIleValLysArgGlyGluL 497
892 .....CAAAACCCCAATGCCGCGACAGGC 915
497 euSerPheAspLysGlnLysTyrAlaGlnAlaAlaGlyAlaAlaGly 513
916 ATAGAGCCGTCACCAAT.....AT 935
514 LeuIleLeValAsnAsnAspGlyThrAlaThrProValThrSerMetAl 530
936 CTTTATGGCAGCCATCCCATCAAGGATTGGAGCTGTCCGGGAAAT 985
530 aLeuThrThrThrPheProThrPheGlyLeuSerSerValThrGlyGlnL 547
986 ACGGCTTGGCGGCATCAGCGCATCTCTCAAGCGGTGCGATGGC 1035
547 ysLeuValAspTrpValThrAlaHisPro.....AspAspSerLeuGly 561
1036 .....GGATCGCATTCGCGAAGGAAATCCGCGCTCAGCGACAA 1076
562 ValLysIleAlaLeuThrLeuValProAsnGlnLysTyrThrGluAspLy 578
1077 TTTTCCGATCGGCATACGCCAATACCGCTCCCTTACCATTCGCGAA 1126
578 smetSerAsp.....PheThrSerTyr.....GlyProValSerAsnLeuS 592
1127 ATATCCGTTCAACATG.....GAG 1146
592 exPheLysProAspIleThrAlaProGlyGlyAsnIleTrpSerThrGln 608
1147 CAGCGTTACGCAAGAAACATC.....ACCTCCTCAACCGTCCGCC 1190
609 AsnAsnAsnGlyTyrThrAsnMetSerGlyThrSerMetAlaSerProPh 625
1191 GTCAAACGGCAAAATGTCAACATGGCAGACCAACGCCCGCAGACAG 1240
```

```
625 eIleAlaGlySerGlnAlaLeuLeuLysGlnAlaLeuAsnAsnLysAsnA 642
1241 CGGTACCGTTT.....GACGGTAAAGGGTTTCG 1269
642 sn...ProPheTyrAlaTyrTyrLysGlnLeuLysGlyThrAlaLeuThr 657
1270 AATTTTGAAGACACGCTGAATATATGACGAAGCTCGATATTCAGAAAT 1319
658 AspPheLeuLysThrValGluMetAsnThrAlaGlnProIleAsnAspI 674
1320 ATCG.....GGGGCGGTATAC 1336
674 eAsnTyrAsnAsnValIleValSerProArgGlnGlnGlyAlaGlyLeu 691
1337 CTAAAGCTAAGCCTGTGTTCATCGCAACCGAGATGGAGTTGATAGG 1386
691 alAspValLysAlaAlaIleAspAla..... 699
1387 AAGCTTAATAATTTGACAACCTCGTGACAGCTGGAGAAAT..... 1428
700 .....LeuGluLysAsnProSerTh 706
1429 .....GTTCAGGAACAGACAGAGAGTCAGTAGTCAGTTTAAAGCCC 1474
706 rValValAlaGlu..... 710
1475 ATGCCCAACGAGATGGGAAATAAACAAGGTTAGATTTTAATCATTTT 1524
711 .....AsnGlyTyrProAlaValGluLeuLysAspPhe 721
1525 ATAGTGTGTATATCAATAAGAAAGGACAGTACAGGAGG.....CA 1568
722 ThrSerThrAspLysThrPheLysLeuThrPheThrAsnSerThrThrHi 738
1569 TAGTCTAACCCCTGGTGATGTACGGGTATACAAACACCTCGGCACCTG 1618
738 sGluLeuThr.....TyrGlnMetAspSerAsnThrA 749
1619 ATAAACATGGGTTTATCAAGCGACAGTGGAA..... 1650
749 spThrAsnAlaValTyrThrSerAlaThrAspProAsnSerGlyValLeu 765
1651 ...ATTAAAGCCTGATGGAAT.....TGGAGGTGAA 1682
766 TyrAspLysLysIleAspGlyAlaAlaIleLysAlaGlySerAsnIleTh 782
1683 ACGAAAAAAGGTGGAAAGTATGACCAACAGCACCATG...TTCCCAA 1729
782 rValProAlaGlyLysThrAlaGlnIleGluPheThrLeuSerLeuProL 799
1730 AAGATTGGGATGAG 1743
799 ysSerPheAspGln 803
seq_name: SwissProt_40:CCAA_MOUSE
seq_documentation_block:
ID CCAA_MOUSE STANDARD; PRT; 2164 AA.
AC P97445;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel 1) (BT)
DE CACNA1A OR CACNA1A4 OR CCH4 OR CAC4 OR CACN3.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT TG LEU-601.
RC STRAIN-DBA/2J;
```

RX MEDLINE=97083572; PubMed=8929530;  
 RA Fletcher C.F., Lutz C.M., O'Sullivan T.N., Shaughnessy J.D. Jr.,  
 RA Hawkes R., Frankel W.N., Copeland N.G., Jenkins N.A.;  
 RT "Absence epilepsy in tottering mutant mice is associated with calcium  
 channel defects".  
 RL Cell 87:607-617(1996).  
 CC -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE  
 ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED  
 IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE  
 CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,  
 CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A  
 GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM  
 CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND  
 ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-  
 IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO  
 DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-  
 GVIA).  
 CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
 LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC; MAINLY FOUND IN THE  
 CEREBELLUM, OLFACTORY BULB, CEREBRAL CORTEX, HIPPOCAMPUS, AND  
 INFERIOR COLLICULUS. IN THE HIPPOCAMPUS, EXPRESSION OCCURS IN  
 PYRAMIDAL AND GRANULE NEURONS, AS WELL AS IN INTERNEURONS.  
 CC PURKINE CELLS CONTAIN PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING  
 A PROMINENT CALCIUM CURRENT IN CEREBELLAR GRANULE CELLS.  
 CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE  
 HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE  
 POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS  
 PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A  
 SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
 CC -!- DISEASE: DEFECTS IN CACNA1A ARE THE CAUSE OF A DELAYED-ONSET,  
 RECESSIVE NEUROLOGICAL DISORDER SEEN IN TOTTERING (TG) MUTANTS,  
 RESULTING IN ATAXIA, MOTOR SEIZURES AND BEHAVIORAL ABSENCE  
 SEIZURES RESEMBLING PETIT MAL EPILEPSY (OR ABSENCE EPILEPSY) IN  
 HUMANS. THERE ARE TWO MORE ALLELES: LEANER (TG(LA)), THAT IS  
 CHARACTERIZED BY SEVERE ATAXIA AND FREQUENT DEATH PAST WEANING.  
 CC BUT NO MOTOR SEIZURES; AND ROLLING NAGAYA (TG(ROL)), THAT PRESENTS  
 AN INTERMEDIARY PHENOTYPE, THE ATAXIA BEING SOMEWHAT MORE SEVERE  
 CC THAT WITH TG, BUT WITHOUT MOTORS SEIZURES. SELECTIVE DEGENERATION  
 OF CEREBELLAR PURKINJE CELLS HAS BEEN SHOWN FOR ALL THESE TYPES OF  
 MUTANTS.  
 CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS  
 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U76716; AAC52940.1; -.  
 DR MGD; MGI:109482; Cacnala.  
 DR InterPro; IPR000637; AT\_hoek.  
 DR InterPro; IPR002077; Ca\_channel.  
 DR InterPro; IPR002111; Cat\_channel\_TrpL.  
 DR InterPro; IPR000636; Cation\_chan\_non\_lig.  
 DR Pfam; PF00520; ion\_trans; 4.  
 DR PRINTS; PR00167; CACHANNEL.  
 DR SMART; SM00384; AT\_hoek; 1.  
 KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
 KW Calcium channel; Glycoprotein; Repeat; Multigene family;  
 KW Calcium-binding; Phosphorylation; Disease mutation.  
 FT REPEAT 17 317 I.  
 FT REPEAT 427 671 II.

FT	REPEAT	1134	1417
FT	REPEAT	1454	1717
FT	DOMAIN	1	52
FT	DOMAIN	53	71
FT	DOMAIN	72	90
FT	DOMAIN	91	108
FT	DOMAIN	109	120
FT	DOMAIN	121	136
FT	DOMAIN	137	144
FT	DOMAIN	145	163
FT	DOMAIN	164	182
FT	DOMAIN	183	202
FT	DOMAIN	203	289
FT	DOMAIN	289	314
FT	DOMAIN	315	441
FT	DOMAIN	442	461
FT	DOMAIN	462	475
FT	DOMAIN	476	495
FT	DOMAIN	496	503
FT	DOMAIN	504	522
FT	DOMAIN	523	532
FT	DOMAIN	533	551
FT	DOMAIN	552	570
FT	DOMAIN	571	590
FT	DOMAIN	591	643
FT	DOMAIN	644	668
FT	DOMAIN	669	1142
FT	DOMAIN	1143	1166
FT	DOMAIN	1167	1183
FT	DOMAIN	1184	1203
FT	DOMAIN	1204	1210
FT	DOMAIN	1211	1234
FT	DOMAIN	1235	1245
FT	DOMAIN	1246	1263
FT	DOMAIN	1264	1282
FT	DOMAIN	1283	1302
FT	DOMAIN	1303	1389
FT	DOMAIN	1390	1414
FT	DOMAIN	1415	1470
FT	DOMAIN	1471	1489
FT	DOMAIN	1490	1503
FT	DOMAIN	1504	1525
FT	DOMAIN	1526	1532
FT	DOMAIN	1533	1552
FT	DOMAIN	1553	1559
FT	DOMAIN	1560	1578
FT	DOMAIN	1579	1597
FT	DOMAIN	1598	1617
FT	DOMAIN	1618	1689
FT	DOMAIN	1690	1715
FT	DOMAIN	1716	2164
FT	DOMAIN	681	686
FT	DOMAIN	1107	1110
FT	DOMAIN	2114	2123
FT	DOMAIN	337	354
FT	SITE	272	272
FT	SITE	622	622
FT	SITE	1363	1363
FT	SITE	1659	1659
FT	MOD_RES	1725	1725
FT	CA_BIND	1743	1754
FT	CARBOHYD	237	237
FT	CARBOHYD	1559	1559
FT	VARIANT	601	601
FT	SEQUENCE	2164 AA;	246978 MW;

E9A80AC5B4758BBC CRC64;

alignment\_scores:







```
FT REPEAT 244 257 1.
FT REPEAT 258 271 2.
FT REPEAT 272 285 3.
FT REPEAT 286 299 4.
FT REPEAT 300 313 5.
FT REPEAT 314 327 6.
FT REPEAT 328 341 7.
FT REPEAT 342 355 8.
FT REPEAT 384 391 9 (INCOMPLETE).
FT REPEAT 392 405 10.
FT REPEAT 406 417 11.
FT REPEAT 418 431 12.
FT REPEAT 432 445 13.
FT REPEAT 446 460 14.
FT REPEAT 461 474 15.
FT REPEAT 475 488 16.
FT REPEAT 489 502 17.
FT REPEAT 503 516 18.
FT REPEAT 517 530 19.
FT REPEAT 907 937 TAU/MAP MOTIF.
FT REPEAT 938 968 TAU/MAP MOTIF.
FT REPEAT 969 1000 TAU/MAP MOTIF.
SQ SEQUENCE 1072 AA; 111914 MW; E8C17A730989F0D2 CRC64;
```

```
alignment_scores:
  Quality: 118.50      Length: 560
  Ratio: 0.459        Gaps: 31
  Percent Similarity: 46.071 Percent Identity: 22.321
```

```
alignment_block:
US-09-303-518D-463 x MAP4_BOVIN ..
```

Align seg 1/1 to: MAP4\_BOVIN from: 1 to: 1072

```
308 CGCCCTTCGACCAACCATCGCTCATTCCGATTCGACGAGCGGTAGT 357
:::|||||::: :::::||||| :::::|||||
600 LysProSerGlnThrSerGluLeuProSerGluThrSerGlyValAl 616
:::|||||::: :::::||||| :::::|||||
358 CCCGTTGACGGATTCACGCTTTACCGCATCCATTGGGACGGATACGAACA 407
: ::::: :::::||||| :::::|||||
616 aLysProGluGluGlyProProThrGlySerValSerGlyAsnAspIleT 633
:::|||||::: :::::||||| :::::|||||
408 CCATCCCGCGCGGCGGTATGACGGCCACAGCGCGGCGGTATCCGCTC 457
||| ::::: :::::||||| :::::|||||
633 hrAlaProPro.....AsnLysGluLeuProProSer 643
:::|||||::: :::::||||| :::::|||||
458 CCAAGCGCGGAGGATATATACAGCTACGACATATAAAGCGG..... 499
||||::: :::::||||| :::::|||||
644 ProGluLysLysThrLysProLeuAlaThrThrGlnProAlaLysThrSe 660
500 .....TGCCCAAAATATCCGCT 518
rThrSerLysAlaLysThrGlnProThrSerLeuProLysGlnThrAla. 676
519 CAACCTGACGACCAACCGGACCGACCAAGCGCTTGCGG..... 559
||||| :::::||||| :::::|||||
677 .....ProThrThrLeuLysGlySerAsnLysLysProMetSerLeu 690
560 ACCGTTTCCACATGCGCGGTATGCTGACGCAAGGAGTAGCGGACGGA 609
::: :::::||||| :::::|||||
691 AlaSerGlySerValProAlaAlaProProLysArgProAlaAlaAlaTh 707
610 TTCAAACGGCCACCGGATACGCGCGGAGCTGGACAGATCGGCAATGC 659
||||| :::::||||| :::::|||||
707 rSerArg.....ProSerThrLeuProSer.....LysAspThrL 719
660 CGCGGAGCTTCAACGGCACTGCAGATA.....TCG 691
||||| :::::||||| :::::|||||
719 ysProLysProValAlaGluAlaLysIleProGluLysArgValSerPro 735
692 TCAAAACATCATCGCGCGGCGGACGAGAA.....TTGTCGCGCGCA 732
||||| :::::||||| :::::|||||
```

```
736 SerLysProAlaSerAlaProAlaValLysProGlySerLysSerThrGl 752
733 GCGATCCGTCGAGGTATAAGCGAAGGCTCAACATTGCTGTCATGCA 782
||||| :::::||||| :::::|||||
752 nAlaValPro.....LysAla.....ProAlaT 760
783 CGGCTTGGGTCTGTTTCCACCAAAACAGATGGCGCGCATCAACGATT 832
||||| :::::||||| :::::|||||
760 hrAlaThrLeuAlaSerProGlySerThrSerArgAsnLeuSerThr... 775
833 TGGCAGATATGGCGCAACTCAAGACTATGCCGACGACGACATCCCGCAT 882
||||| :::::||||| :::::|||||
776 .....ProLeuProLysArgProThrAlaIl 784
883 TGGCAGTCCAAACCCCAATGCCGACAAAG.....GCATAGAACCGT 926
| :::::||||| :::::|||||
784 eLysThrGluGlyLysProAlaGluIleLysLysMetAlaThrLysSerA 801
927 CAGCAATATCTTTATGGCAGCCATCCCATCAAGGGATTGGAGTGTCC 976
::: :::::||||| :::::|||||
801 laProAlaAspLeuSerArgProLysSerThrThrSerSerValLys 817
977 GGGAAATACGGCTTGGCGGCATCACGGCACATCTGTCAAGCGGTG 1026
::: :::::||||| :::::|||||
818 LysSerThrThrValProGlyThrAla..... 826
1027 CAGATGGCGCGCATCGCATTCGGAAGGAAATCCGCGTCAGCGACAA 1076
||||| :::::||||| :::::|||||
827 .....ProProAlaGlyAla. 831
1077 TTTTCCGATCGGCATACGCCCAATACC.....CGTCCCTTACC 1117
||| ||| :::::||||| :::::|||||
832 .....ProSerArgAlaArgProThrAlaThrProProArgProSerGly 846
1118 ATTCCGAAATATCGTTCAACTTGGAGCAGCGTTACGCAAAAGAAAC 1167
||| :::::||||| :::::|||||
847 ThrProProValAspLysLys.....ProThrAlaAlaLysProTh 860
1168 ATCACTCTCTCAACGTCGCCCGCTCAACCGGCAAAATGTCAAACTGGC 1217
||||| :::::||||| :::::|||||
860 rSerSerAlaProArgLeuGlyArgValAlaAla..AsnAlaSerAlaPr 876
1218 AGACAACCCACCGAAGACAGCGGTACCGTTTGACGGTAAAGGGTTTC 1267
||| :::::||||| :::::|||||
876 oAspLeuLysAsnValArgSerLysVal.....GlySerT 888
1268 CCAATTTTGAAGCAGCTGAAATATGATACGAAGCTGATATTCAAGAA 1317
||| :::::||||| :::::|||||
888 hrGluAsnIleLysHisGlnProGlyGlyArgAlaLysValGluLys 904
1318 TTATCGGGGGCGGTATACCTAAGCTAAGCTGTGTTGATGCG..... 1362
||| :::::||||| :::::|||||
905 LysThrGluAlaAlaProAlaArgLysProGluProAsnAlaValTh 921
1363 .....AAACCG.....AGAT 1372
921 rLysAlaAlaGlyProIleGlyAsnAlaGlnLysProProThrGlyLysV 938
1373 GGGAGGTGTATAGGAAG.....CTTAATAATTGACAACCTCGTGAG 1413
||| :::::||||| :::::|||||
938 alGlnIleGlnAsnLysLysValAspLysValSerLysValSerLysCys 954
1414 CAGTGGAGAAAAATGTTGAG.....GAAACGAGGAAG 1445
||| :::::||||| :::::|||||
955 GlySerLysAlaAsnIleLysHisLysProGlyGlyGlyAspValLysIl 971
1446 AAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGGCAACGAGAAATGGGAA 1495
::: :::::||||| :::::|||||
971 eGluSerGlnLysLeuAsnPhelLysGluLysAlaGlnAla..... 984
1496 ATAAACAGGG...TTAGAT...TTTAATCATTTT...ATAGGTGGTGTAT 1536
||||| :::::||||| :::::|||||
985 ..LysValGlySerLeuAspAsnValGlyHisLeuProAlaGlyGlyAla 1000
```

```

271 ATTGTCGGCTTTTCCGAT.....CACGGGCACAAATTCCATTCG... 309
::: :::::||||| ::::: ||||| ::::: ||||| ::::: |||||
242 ValGluLysPheThrAlaLysHisGlyArgTyrPheAsnSerLy 258
::: :::::||||| ::::: ||||| ::::: ||||| ::::: |||||
310 .....CCCTTCGAC...AACCATGCGCTCACATTCGGAT.....TCTGACG 346
||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
258 sValProTyrGlyPheAsnTyrAlaAspAsnAsnAspThrIleThrAspA 275
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
347 AAGCCGGTAGTCCCGTGTACGAGGATTCAGCGCTTTACCGCATCCATTGGGAC 396
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
275 spThr.....ValAsp..... 278
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
397 GGATAGCAACACCATCCCGCCGCGAGCGCTATGACGGGCCACGGCGCGG 446
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
279 .....GluGlnHis.....GlyMetHisValAlaGlu 287
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
447 CTATCCCGCTCCCAAGAGCGGAGGGATATATACAGCTACGACATATAAG 496
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
287 YlleIleGlyAlaAsnGlyThrGlyAspAspProAlaLysSerValG 304
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
497 GCGTGTGCCCAAAATATCCGCCTC.....AACCTG 525
||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
304 lyValAlaProGluAlaGlnLeuLeuAlaMetLysValPheThrAsnSer 320
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
526 ACCGACACACCGCAGCACCGGA.....CAACGCGCTGCCGA 560
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
321 AspThrSerAlaThrThrGlySerAlaThrLeuValSerAlaIleGluAs 337
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
561 CCGTTTCCCAATCGCGGCGCTATGCTGACGCAAGCAGTAGCGCAGGAT 610
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
337 pSerAlaLysIleGlyAlaAspValAlaAsnMetSerLeuGlySerAspS 354
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
611 TCAACCGCGCACCGCATACAGCCCGCGAGCTGCGACAGATCGGGCGCAATGCC 660
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
354 ArgGlyAspGlyThrLeuGluAsnProGluLeuAlaAlaValGlnAspAla 370
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||

```

```

1537 ATCAATAAGAAAGGCACAGTAACAGGAGCGATAGTCTAAACCCGTGGTGA 1586
      :::::  ::::|  |||||
1001 ValLysThrGluGly.....GlyGly..... 1007
      :::::  ::::|  |||||
1587 TGTACGGGTGATACAAACAACCTCGGCACCTGATAAACATGGGTTTATC 1636
      |||  |||||  :::::  |||||
1008 .....SerGluAlaProProCysProGlyPro...P 1017
      :::::  ::::|  |||||
1637 AAGCAGACAGTGGAAATTAAGAGCCTGATGGAAGTTGGAGGTGAAACG 1686
      |||  |||||  :::::  |||||
1017 roAlaGlyCylGluLeuAlaIleProGluAlaAlaProGluAlaGlyAla 1033
      :::::  ::::|  |||||
1687 AAAAAGGTGGAAAGTAGTACCAAGACACACCATGTTCCCAAAGATTG 1736
      :::::  ::::|  |||||
1034 ProAlaSerAlaSerGlyLeuSerGlyHisThrThrLeuAlaGlyCylG 1050
      |||||  :::::  |||||
1737 GGATGAGGCTAGAAATAGGCTGCAAGTTACTTCGGCTGGGAAAGTAGAA 1786
      |||||  :::::  |||||
1050 yspGlnArgGlu.....AlaGlnThrLeuAspSerGlnI 1062
      :::::  ::::|  |||||
1787 TAATGCTTAAGGATAATAAATGCGAGGT 1815
      ||||:  :::::  |||||
1062 leGln...GluThrSerLysTrpLeuGly 1070

seq_name: SwissProt_40:P2p_LACLC
seq_documentation_block:
P2p_LACLC STANDARD: PRT; 1902 AA.
AAC P15293;
01-APR-1990 (Rel. 14, Created)
01-APR-1990 (Rel. 14, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
P1I-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
associated serine proteinase) (LP151).
PRT.
Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
Plasmid pLP763.
OG Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=89313288; PubMed=2501630;
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
from Streptococcus lactis NCDO763.";
RL Mol. Microbiol. 3:359-369(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the p1 and p4 positions, and
CC pro in the p2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X14130; CAA32350.1; -.
DR PIR; S06997; S06997.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.019; -.

```





RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;  
RX MEDLINE=91288516; PubMed=1648226;  
RA Starr T.V.B., Prystay W., Snutch T.P.;  
RT "Primary structure of a calcium channel that is highly expressed in  
RL the rat cerebellum";  
RN proc. Natl. Acad. Sci. U.S.A. 88:5621-5625(1991).  
RP [2].  
RX SEQUENCE OF 1435-1667 FROM N.A. (CLONE RBA-65).  
RA MEDLINE=90239020; PubMed=1692134;  
RT Snutch T.P., Leonard J.P., Gilbert M.M., Lester H.A., Davidson N.;  
RL "Rat brain expresses a heterogeneous family of calcium channels";  
RN proc. Natl. Acad. Sci. U.S.A. 87:3391-3395(1990).  
RP [3].  
RX SEQUENCE OF 1354-1659 FROM N.A. (CLONE RKCB).  
RC TISSUE=Kidney;  
RX MEDLINE=93066265; PubMed=1279681;  
RA Yu A.S.L., Hebert S.C., Brenner B.M., Lytton J.;  
RT "Molecular characterization and nephron distribution of a family of  
RL transcripts encoding the pore-forming subunit of Ca2+ channels in the  
RC kidney";  
RX proc. Natl. Acad. Sci. U.S.A. 89:10494-10498(1992).  
CC -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE  
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED  
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE  
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,  
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A  
CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM  
CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND  
CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-  
CC IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO  
CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-  
CC GVIA).  
CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: IN THE BRAIN, AT LEAST FOUR ISOFORMS ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC; PURKINJE CELLS CONTAIN  
CC PREDOMINANTLY P-TYPE VSCC. THE Q-TYPE BEING A PROMINENT CALCIUM  
CC CURRENT IN CEREELLAR GRANULE CELLS. ALSO FOUND IN HEART, IN  
CC KIDNEY DISTAL CONVOLUTED TUBULE (DCT), AND IN PITUITARY.  
CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE  
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE  
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS  
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A  
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS  
CC FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: M64373; AAA40806.1; -;  
CC EMBL: M9222; AAA40896.1; -;  
CC InterPro: IPR000637; AT\_hook.  
CC InterPro: IPR002077; Ca\_channel.  
CC InterPro: IPR002111; Cat\_channel\_TrpL.  
CC InterPro: IPR000636; Cation\_chan\_non\_lig.  
CC InterPro: IPR001682; Channel\_pore\_Ca\_Na.  
CC Pfam: PF00520; Ion\_trans; 4.  
CC PRINTS: PR00167; CACHANNEL.  
CC SMART: SM00384; AT\_hook; 1.  
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW

KW Calcium channel; Glycoprotein; Repeat; Multigene family;  
KW Calcium-binding; Phosphorylation; Alternative splicing.  
FT REPEAT 365 I.  
FT REPEAT 475 II.  
FT REPEAT 1182 1182 III.  
FT REPEAT 1502 1502 IV.  
FT DOMAIN 1 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 101 S1 OF REPEAT I (POTENTIAL).  
FT DOMAIN 120 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 139 S2 OF REPEAT I (POTENTIAL).  
FT DOMAIN 157 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 169 S3 OF REPEAT I (POTENTIAL).  
FT DOMAIN 185 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 193 S4 OF REPEAT I (POTENTIAL).  
FT DOMAIN 212 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 231 S5 OF REPEAT I (POTENTIAL).  
FT DOMAIN 251 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 338 S6 OF REPEAT I (POTENTIAL).  
FT DOMAIN 363 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 490 S1 OF REPEAT II (POTENTIAL).  
FT DOMAIN 510 S2 OF REPEAT II (POTENTIAL).  
FT TRANSMEM 524 S3 OF REPEAT II (POTENTIAL).  
FT DOMAIN 544 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 552 S3 OF REPEAT II (POTENTIAL).  
FT DOMAIN 571 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 581 S4 OF REPEAT II (POTENTIAL).  
FT DOMAIN 600 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 618 S5 OF REPEAT II (POTENTIAL).  
FT DOMAIN 639 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 692 S6 OF REPEAT II (POTENTIAL).  
FT DOMAIN 716 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1191 S1 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1214 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1231 S2 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1251 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1252 S3 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1282 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1283 S4 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1311 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1330 S5 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1350 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1351 S6 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1437 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1438 S1 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1463 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1519 S2 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1538 S3 OF REPEAT IV (POTENTIAL).  
FT TRANSMEM 1552 S4 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1574 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1581 S5 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1601 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1608 S4 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1626 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1645 S5 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1665 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1666 S6 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1737 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1738 S1 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1764 S2 OF REPEAT IV (POTENTIAL).  
FT TRANSMEM 2212 POLY-GLY.  
FT DOMAIN 13 POLY-GLU.  
FT TRANSMEM 729 POLY-HIS.  
FT DOMAIN 734 BINDING TO THE BETA SUBUNIT (BY  
FT TRANSMEM 1155 SIMILARITY).  
FT DOMAIN 1518 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT TRANSMEM 2162 (BY SIMILARITY).  
FT DOMAIN 2171 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT TRANSMEM 385 (BY SIMILARITY).  
FT TRANSMEM 402 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT SITE 320 (BY SIMILARITY).  
FT SITE 320 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT SITE 670 (BY SIMILARITY).  
FT SITE 1411 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT SITE 1411 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT SITE 1707 (BY SIMILARITY).  
FT MOD\_RES 1773 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
FT CA\_BIND 1791 BY SIMILARITY.  
FT CARBOHYD 285 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1607 N-LINKED (GLCNAC. .) (POTENTIAL).

```

FT   VARSPLIC 1602 1602      G -> GNP (IN ISOFORM RKC8) .
SQ   SEQUENCE 2212 AA; 251525 MW; DFB7D7A0F553AFA52 CRC64;

alignment_scores:
    Quality: 118.50      Length: 461
    Ratio: 0.620        Gaps: 27
    Percent Similarity: 41.432      Percent Identity: 23.861

alignment_block:
US-09-303-518D-463 x CCAA_RAT      ..
Align seg 1/1 to: CCAA_RAT from: 1 to: 2212

183 CGGCCATATCGGATTTGGGAACATACA.....AAGCCCATCAGTTGG 223
||||| ||| ||||| |||||
810 ArgProLeuArgProAspValThyHisLeuAspArgProLeuValVa 826
224 GCCACCTGATGATTCAACAGCGCGCGTGAAGGAAATATCGGCTACATT 273
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
826 lAspProGlnGluAsnArgAsnAsnThrAsnLys..... 838
274 GTCCGCTTTCCGATCAGCGGGCAAAATTCATTCCGCTTCGACAAACCA 323
||||| ||| ||||| |||||
839 .....SerArgAla.....ProGluAlaLeuArgGlnThr 848
324 TGCTCATCATTCGGATTCTGCAGCAGCCGGTAGTCCGCTTGACGGATTCA 373
: : : : : : : : : : : : : : : : : : : : : : : : : :
849 Ala..... 849
374 GCCTTTACGCGATCCATTGGGAGCGATAGAACACACCCGCCGACGCGC 423
: : : : : ||| ||||| |||||
850 .....ArgProArgGluSerAlaArgAspProAspAlaArgAlaArg 863
424 TA.....TGACGGGCCACAGGCGG..... 443
863 latrpProSerSerProGluArgAlaProGlyArgGluGlyProTyrGly 879
444 CGGCTATCCGCTCCCAAGGCGCGAG..... 470
||| ||||: ||||: ||||: ||||:
880 ArgGluSerGluProGlnArgGluHisAlaProProArgGluHisVa 896
471 .....GGATATATACAGCTAC 486
896 lProTTrpAspAlaAspProGluArgAlaLysAlaGlyAspAlaProArg 913
487 GACATAAAGCGGTTGCCAAATATCCGCTCAACCTGACCGACGACAAACG 536
||||| : : : : : ||| : : : : :
913 rghisThrHisArg.....ProValAlaGluGlyGluPro 924
537 CAGCACCGGACAAAGCGCTTGCAGCGCTTCCCAATCCGCGCGCTATGC 586
: : : : : : : : : : ||||| : : :
925 ArgArgHisArgAlaArgArgProGlyAspGlu.....Pr 937
587 TGACGCAAGGATAGCGCAGGATCAACACGCGCCACCCGATACAGCCC. 635
||| ||| ||||| ||||| : : : ||| : : : |||
937 oAspAspArgProGluArgArgProArgProArgAspAlaThrArgProA 954
636 ..CGAGCT.....GGACAGATCGGGCAATCGCCGCAAGCGCTT 671
||||| ||||| : : : : : : : : : : : : : : : : : : : :
954 laArgAlaAlaAspGlyGluGlyAspAspGlyGluArgLysArg..... 968
672 CAACGCGCTGCAGATAT.....CGTCNAAAACA 700
||||| ||||| : : : : : ||||| : : : : : : : : : : : : : :
969 ...ArgHis...ArgHisGlyProProAlaHisAspAspArgGluArgAr 983
701 TCATCGCGCGCGCAGAGAAATTTGCGCGCAGCGCATCGCTGCAGGGT 750
: : : : : ||||| : : : : : ||||| : : : : :
983 gHisArgArgLys.GluSerGlnGlySerGlyValProMetSerGly 999
751 ATAACGGAAGGCTCAACATGTCTCATGCACGCGTGGGTGCTGCTTC 800
: : : : : : : : : : : : : : : : |||||
1000 ProAsnLeuSerThrThrArgProIleGlnGlnAspLeuGly..... 1013

```